

**INVESTIGATING THE RELATIONSHIP BETWEEN STUDENTS' SELF-
ASSESSMENT AND RATINGS OBTAINED FROM A FORMAL DLPT5
READING SKILL**

Doctoral Dissertation Research

Submitted to the
Faculty of Argosy University, Bay Area Campus
College of Education

In Partial Fulfillment of
the Requirements for the Degree of

Doctor of Education

by

Mohamad Ahmed Alkhatatbeh

March 2014

**INVESTIGATING THE RELATIONSHIP BETWEEN STUDENTS' SELF-
ASSESSMENT AND RATINGS OBTAINED FROM A FORMAL DLPT5
READING SKILL**

Copyright ©2014

Mohamad Ahmed Alkhatatbeh

All rights reserved

**INVESTIGATING THE RELATIONSHIP BETWEEN STUDENTS' SELF-
ASSESSMENT AND RATINGS OBTAINED FROM A FORMAL DLPT5
READING SKILL**

A Dissertation

Submitted to the
Faculty of Argosy University, Bay Area Campus
College of Education

In Partial Fulfillment of
the Requirements for the Degree of

Doctor of Education

by

Mohamad Ahmed Alkhatatbeh

Argosy University

March 2014

Dissertation Committee Approval:

Committee Chair: Dr. Quamina Afriye

Date

Program Chair: Dr. Dennis Frese

**INVESTIGATING THE RELATIONSHIP BETWEEN STUDENTS' SELF-
ASSESSMENT AND RATINGS OBTAINED FROM A FORMAL DLPT5
READING SKILL**

Abstract of Doctoral Dissertation Research

Submitted to the
Faculty of Argosy University, Bay Area Campus
College of Education

In partial Fulfillment of
the Requirements for the Degree of

Doctor of Education

by

Mohamad Ahmed Alkhatatbeh

Argosy University

March, 2014

Dr. Quamina Afriye

Dr. Lespier Mary

Department: College of Education

ABSTRACT

The purpose of the study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). To conduct this study, the researcher investigated the correlation between the two assessments: ratings obtained from the can-do-scale (self-assessment instrument survey) of Arabic reading ability and scores obtained from the valid and reliable DLPT5-R test of Arabic reading ability. This study used the quantitative correlational method to determine the validity and reliability of the new self-assessment testing instrument of Arabic reading ability. The participants of the study included 107 U.S. male and female military students from the four branches of service: Army, Navy, Air force, and Marines. The participants are also studying the Arabic language for 63 weeks at the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, California. The results of the Spearman's rho correlation analysis showed that there is no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability. The results of the Chi-square test showed that there is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability. The results of the ordinal logistic regression showed that there is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

ACKNOWLEDGMENTS

I would like to gratefully and sincerely thank my dissertation committee supervisor Dr. Quamina Afriye who kept encouraging me in every step of my research and since I started my first year at Argosy University; writing this doctoral thesis would not have been possible without his guidance, experience, knowledge, understanding, support, and friendship; Dr. Quamina is a great advisor.

I would also like to recognize and thank the dedicated member of my dissertation committee member, Dr. Lespier Mary for all her time, kindness, wisdom, and input in this study. I would also like to thank Dr. Griffith Scott and all Argosy faculty members and leadership for all their support, guidance, and expertise.

I would also like to recognize and express my deepest gratitude to Dr. Jackson Gordon, Dr. Rogan Seumas, Dr. Elfiky Salem, and Dr. Boussalhi Abdelfattah for all their advice, knowledge, support, and friendship. I would also like to thank all the countless people, friends and colleagues who shared their wisdom and kindness in this endeavor, and regret not being able to mention everyone here by name, but exceptional thank to every one of you.

DEDICATION

I would like to gratefully thank and dedicate this dissertation to my parents, Abu Farouq and Um Farouq, my brothers, my sisters, my daughters, my father-in-law and mother-in-law, all of my extended relatives, and all my friends for their unending encouragement, support, motivation, and inspiration in obtaining my doctoral degree. I would also like to dedicate this dissertation to my wife, Um Sajida, who shared the same struggle with me and who was always encouraging and supporting me through every step of this thesis.

TABLE OF CONTENTS

	Page
TABLE OF TABLES	vi
TABLE OF FIGURES	viii
TABLE OF APPENDICES	ix
CHAPTER ONE: THE PROBLEM	1
Background of the Problem	1
Problem Statement	3
Purpose of the study	7
Significance of the Study	7
The Defense Language Proficiency Test 5 Reading (DLPT5) Testing System	8
DLPT5 Test Content	8
DLPT5 Test Design	9
DLPT5 Reliability and Validity	9
Theoretical Foundation	10
Constructivist Learning Theory	10
Multiple Intelligences Theory	11
Social Cognitive Theory	13
Research Questions and Hypothesis	14
Definition of Terms	15
Limitations and Delimitations of the Study	16
Limitations	16
Delimitations	16
CHAPTER TWO: REVIEW OF THE LITERATURE	18
Introduction	18
Theoretical Foundation	18
The Social Cognitive Learning Theory	18
The Constructive Learning Theory	20
Multiple Intelligences Learning Theory	21
Current Literature Related To Study	22
Self-Evaluation	23
The Importance of Self-Assessment	23
Language Use in Self-Assessment (Methodology)	25
Students Self-Placement	26
Perceptions of Self-Assessment	26
Self-Assessment Requirements	27
Distinction Between Self-Assessment and Self-Evaluation	27
Reasons to Use Self-Assessments	28
Self-Assessment Definition	28
Self-Assessment Methods	29
Assessment Categorization	29
Need for Self-Assessment	29

Self-Assessment Reliability and Validity	30
Steps to Assure Reliability of Self-Assessment.....	32
Literature Research Review	33
Self-Assessment Process.....	34
Self-Assessment Implications for Practice	34
Self-Assessment Studies	35
Motivational Theory (Maslow's Hierarchy of Needs).....	35
Reading Assessment	36
Informal Assessment.....	37
Training in Using Self-Assessment	38
Self-Assessment Verses Other Tests	39
Justifications for Using Self-Assessment.....	39
Summary	40
CHAPTER THREE: METHODOLOGY	42
Introduction.....	42
Research Method and Design	42
Participants.....	45
Operationalization of the Variables	46
Instrumentation	47
Validity and Reliability.....	49
Rules for Scoring Self-Assessment.....	52
Data Collection	55
Data Analysis	55
Protection of Human Participants	59
Summary	60
CHAPTER FOUR: FINDINGS	61
Results of Descriptive Statistics.....	62
Correlation Results for the Arabic CDS and the Arabic DLPT5-R Test.....	75
Chi-Square Test Results.....	78
Results of Ordinal Logistic Regression	80
Summary	83
CHAPTER FIVE: DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS ...	84
Introduction.....	84
Overview of the Study	84
Summary of the Results	86
Discussion of the Results in Relation to Literature	87
Limitations	91
Implication of the Results for Practice.....	92
Recommendations for Further Research.....	93
Conclusion	94
REFERENCES	96

TABLE OF TABLES

Table	Page
1. Survey Items Scoring Rules.....	53
2. CDS Scoring Rules	54
3. Arabic CDS Scores	63
4. Arabic DLPT5-R Scores	64
5. Highest Education Level Completed	65
6. Military Branch.....	66
7. Military Rank	67
8. Gender.....	67
9. Age.....	68
10. Previous Experience Reading a Language Other Than English	69
11. Previous Language Reading Experience Other Than English.....	70
12. Age First Started Reading English.....	70
13. Previous Experience Studying Arabic Before DLIFLC	71
14. Length of Time Studying Arabic Before DLIFLC	72
15. Previous Experience Studying Other Foreign Languages Before DLIFLC	72
16. Previously Studied Other Foreign Language Before DLIFLC	74
17. Length of Time Studying Other Foreign Language Before DLIFLC.....	75
18. Spearman's Correlation Result between Arabic CDS Scores and Arabic DLPT5-R.	76
19. Intraclass Correlation Coefficient Results	78
20. Cross Tabulation Results Between Arabic Reading Scores	79
21. Chi-Square Test Results.....	80

22. Results of Ordinal Logistic Regression Test	82
---	----

TABLE OF FIGURES

Figures	Page
1. Arabic CDS Scores	63
2. Arabic DLPT5-R Scores	64
3. Highest Education Level Completed	66
4. Military Branch	66
5. Military Rank	67
6. Gender	68
7. Age	68
8. Previous Experience Reading a Language Other Than English	69
9. Age First Started Reading English	70
10. Previous Experience Studying Arabic Before DLIFLC	71
11. Length of Time Studying Arabic Before DLIFLC	72
12. Previous Experience Studying Other Foreign Languages Before DLIFLC	73
13. Scatter Plot of Arabic CDS Scores and Arabic DLPT5 Reading Scores	76

TABLE OF APPENDICES

Appendix	Page
A: DLPT5-Reading Multiple Choice Format	103
B: DLPT5-Reading Constructed-Response Format.....	105
C: Interagency Language Roundtable Language Skill Level Descriptions	107
D: Self-Assessment Survey of Reading Proficiency	112
E: CDS Instrument.....	114
F: Informed Consent Form.....	121
G: The Approval Letter to Conduct the Study at DLIFLC	124

CHAPTER ONE: THE PROBLEM

Background of the Problem

The Defense Language Institute Foreign Language Center (DLIFLC) provides foreign language education, training, evaluation, and sustainment for Department of Defense (DoD) personnel in order to ensure the success of the Defense Foreign Language Program and enhance the security of the nation. To achieve the ultimate goals, vision, and mission of the DLIFLC, there should be more studies that explore learning techniques, teaching strategies and self-assessments of foreign language learning.

The DLIFLC (2014) mission statement is to, "provide culturally-based foreign language education, training, evaluation, research, and sustainment for DoD personnel in order to ensure the success of the Defense Language Program and enhance the security of the nation" (para. 1). The DLIFLC (2014) vision statement is "delivering the world's best culturally-based foreign language training and education – at the point of need" (para. 2). O'Connell & Norwood (2007) highlighted that,

the National Security Language Initiative (NSLI) announced that we must increase the nation's capacity to provide experts with critical language skills - in languages such as Arabic, Chinese, Farsi/Dari, Hindi/Urdu and Turkic - determined to be vital to national security and foreign policy. (pp. 4-5)

Ochoa (2012) indicated that to increase the global competencies of all U.S. students, we need to include foreign language competency. There has been a dramatic decrease in the number of U.S. students enrolled in foreign language studies over the past decades. Ochoa (2012) also noted, only 30% of U.S. students studying foreign languages at the secondary level, and only 8% enrollment at the postsecondary level. This is a low percentage of students studying foreign languages compared with the European Union (EU) member states, where 59.6% of students at International Standard of Classification

(ISCED) Level 3 are studying two or more languages (EuroStat, 2012). ISCED Level 3 is equivalent to upper secondary education where students are aged 15-16.

Nordin (2012) stressed that “human skills in foreign languages, knowledge of cultures, and expertise in regions all play a key role in or directly support all foreign intelligence disciplines” (p. 1). The author further states that intelligence analysts, without knowing other languages, depend on others’ interpretations and translations, which may lead to errors and unintentional bias when they transfer the meaning.

Former Defense Secretary Panetta (2011) believes in the importance of having a strong language ability to secure and protect the U.S. nation. Panetta stated, “language, regional, and cultural skills are enduring war fighting competencies that are critical to mission readiness in today’s dynamic global environment” (p. 1). Former Defense Secretary Panetta (2011) also noted the importance of teaching foreign languages and cultures to U.S. citizens,

Commanders must ensure that deploying units, leaders, and staff receive the language and cultural training that is commensurate with their missions and responsibilities. We must also increase and sustain the foreign language proficiency of our language and regional professionals if we are to be able to understand and plan for the future mission. (p. 1)

Goodman (2012) stated, “For the United States, knowledge of foreign languages and cultures is essential to our national security and to preparing Americans to meet the demands of the global workforce” (p. 2). Goodman (2012) believes that U.S. higher education should make foreign language proficiency a requirement for graduation, as it was for all colleges and universities 100 years ago. Goodman (2012) mention that in the years 2009 and 2010 in the United States there were only 121 students graduating with a

Bachelor's degree in the Arabic language, and there were only 456 students who graduated with a Bachelor's degree in Chinese.

Fenstermacher (2012) stated, "Every sector of the U.S. economy depends on language services for revenue, profit growth, job creation, product innovation, and research and development" (p. 1). Thomas-Greenfield (2012) said that "The United States and the world face great perils and urgent foreign policy challenges, including regional conflicts, wars, the global economic crisis, weapons of mass destruction, climate change, worldwide poverty, food insecurity, pandemic disease, and terrorism" (p. 1).

According to Richards (2001), in any institution, teachers may vary in language proficiency, teaching experience, skills, teaching style, beliefs, and principles. In planning a cultural program for students who are learning a second language, it is therefore important to know the kinds of teachers the program will depend on to ensure the success of the program mission. In addition, in Richards (2001) book *Curriculum Development in Language Teaching*, he discusses the teacher's behavior, the student's expectations for the roles of teachers, and the importance of having an extensive orientation to the teacher's teaching context.

Problem Statement

The United States is suffering from a lack of knowledge about foreign languages and cultures, which threatens the U.S. security and economy (O'Connell & Norwood, 2007). O'Connell and Norwood (2007) mentioned that in 2007 the National Research Council reported that "A pervasive lack of knowledge about foreign cultures and foreign languages threatens the security of the United States as well as its ability to compete in the global marketplace and produce an informed citizenry" (p. 1). O'Connell and

Norwood (2007) further stated, “Current efforts to develop language assessments and to effectively apply developments in technology to language assessment and the support of language instruction suffer from a dispersion of resources” (p. 4). Taha (2007) noted that Former President George W. Bush emphasized the need to increase the number of Americans who learn foreign languages. He identified Arabic, Hindi, Russian, Farsi, and Chinese as critical needs, and discussed that there have not been detailed studies on Arabic language learning or teaching methods in the U.S.

The limited scholarly research about learning language skills is considered a threat to the United States; learning foreign languages becomes critical because the world is facing new challenges in many fields, including deteriorating foreign language competencies; “The entire world is faced with new challenges in developing/consolidating global understanding, intercultural communication, peace and economic prosperity” (Taha, 2007, p. 13).

The September 11, 2001 tragic attack on the World Trade Center in New York revealed the language shortfalls of the United States (Akaka, 2012). The 9/11 Commission was concerned about the shortage of personnel who were knowledgeable and proficient in the Middle Eastern languages at the Central Intelligence Agency, the Federal Bureau of Investigation, the Department of Defense, the Department of Homeland Security, and the Department of State.

These shortages of personnel hinder understanding of the threats that are facing the United States. Akaka (2012) noted, “Because of these shortages, agencies are forced to fill language-designated positions with employees that do not have those skills. Agencies then have to spend extra time and funds training employees in these languages”

(p. 1). To improve U.S. diplomatic readiness to successfully deal with the challenges of U.S. security and economic success, the U.S. needs to develop a foreign language strategy and capacity, in order to build the Federal government's foreign language skill.

After the September 11 terrorist attacks, the U.S. launched a global war on terror which required U.S. soldiers to communicate effectively to accomplish their mission and to be able to save soldiers' lives in the United States and when they travel overseas.

Husseinali (2006) found that after September 11, 2001, the number of students who enrolled in Arabic as a foreign language (AFL) doubled and is expected to keep increasing. Although the DoD is committed to finding the most capable force for deployment and the language capability and proficiency that are a significant and critical component of that same deployed force, the DoD is facing challenges in training capable personnel to deal with present and projected operational needs.

Junor (2012) states that the Department of Defense (DoD) lacks personnel who have the required language proficiency level to effectively do the responsibilities required of them. Right now, the personnel with the required language proficiency level is only at 28% (10,377) out of a total of 36,983 military language positions. These military language positions are currently filled with personnel who have the required language proficiency level as these positions are identified as having language requirements. The remaining positions may be filled with personnel who do not have the required language proficiency level.

If all positions are filled with qualified personnel, the DoD will meet their requirements to strengthen relationships with existing allies, remain engaged in the international arena, and continue communicating with local people and their senior

officials. If the language training objectives are met, then as a result the U.S. security and the security of the global partners will be safe and protected.

The required language proficiency level for the DLIFLC students to pass the DLPT5 test is obtaining at least 2 in the reading skill, 2 in the listening skill, and 1+ in the speaking skill based on the Interagency Language Roundtable (ILR) skill level descriptions. The *National Security Workforce* Flagship language program (such as the Pilot African Language Initiative, Boren Scholarships, and Fellowships) is designed to increase the pool of experts with critical language proficiency and regional expertise, and bring students to ILR Level 3, or general Professional proficiency.

This study seeks to address the lack of personnel with the required language proficiency level (of 2 in the reading skill, 2 the listening skill, and 1+ in the speaking skill) for military language positions at the DoD. There are no known studies identifying the relationship between students' Arabic reading self-assessment and ratings obtained from a formal Arabic Defense Language Proficiency Test 5 (DLPT5) in the reading skill at the Defense Language Institute (DLI) (Jackson, 2012).

This study aims to address such problem (lack of personnel with the required language proficiency level) through developing a self-assessment instrument for Arabic reading ability that teachers, students and U.S. soldiers at the DLIFLC can use to monitor and improve their required language proficiency level. This study seeks to investigate whether a self-assessment instrument for Arabic reading would provide an accurate estimation of the Arabic DLPT5 reading test results.

Purpose of the study

The purpose of the study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). To accomplish the study, the researcher investigated the correlation between the two assessments: ratings obtained from the CDS (a self-assessment instrument) and scores obtained from the valid and reliable Arabic DLPT5-R Test of Reading ability. The validated CDS instrument was used to measure students' self-assessment and the control variables of the highest education level completed, military branch, military rank, gender, age, and previous experience with language learning were included in the analysis. These control variables can provide further understanding of the effects of demographic variables on self-assessment and language proficiency.

Significance of the Study

The research results of this study could provide valuable information related to testing students and the evaluation of their language proficiency. Teachers and academic specialists will be able to monitor students' success from the first day of class until they graduate. The Can-Do-Scale (CDS) self-assessment instrument will measure the Interagency Language Roundtable (ILR) Skill Level Descriptions of Reading skill.

DLIFLC students will be able to use the CDS instrument of Arabic reading ability to monitor their progress and possibly improve their learning since they can be transformed into active learners who take control of their learning and seek ways to achieve their goals.

Teachers at the DLIFLC will be able to use the self-evaluation instrument of Arabic reading ability to monitor students' current level of proficiency, provide feedback, and create action plans to increase students' language proficiency in Arabic Reading comprehension. Woolfolk (2007) emphasized that "self-evaluation can accompany self-correction. Students first evaluate, then alter and improve their work, and finally, compare the improvement to the standards again" (p. 235).

The Defense Language Proficiency Test 5 Reading (DLPT5) Testing System

The DLPT5-Reading test assesses the language proficiency of native speakers of English (U.S. civilians and military personnel) who have learned a foreign language as a second language, and regardless of how such foreign language has been obtained or been taught. The DLPT5 questions are designed to measure student proficiency according to the ILR Skill Level Descriptions from level 0+ to level 4. All DLPT5s are administered through a computer and in multiple-choice (MC) format. The DLPT5 results are used to make decisions about incentive pay, operational readiness, and training and assignments for military personnel or civilians with language experience and skill working in the United States government.

DLPT5 Test Content

The DLPT5 aims to evaluate the general language proficiency in reading and listening of examinees who are native English speakers and who have learned a foreign language as a second language, and regardless of how the foreign language has been acquired. The test content is not tied to any certain language program because the proficiency orientation of the test is broad.

The DLPT5-Reading passages are obtained from authentic real life sources as much as possible. These resources include internet articles, newspapers, and magazines. The content that the DLPT5-R test includes is topics in geography, economics, military, security, social, culture, politics, science, and technology. The DLPT5-Reading test assesses the examinee's abilities in finding information, reading to recognize and comprehend main ideas transmitted by the writer.

DLPT5 Test Design

The DLPT5s for reading exists in many languages and include both lower-level tests and upper-level tests. The lower-level tests measure the ILR reading proficiency levels from 0+ - 3, while the upper-level tests measure the ILR reading proficiency levels 3+ and 4. Examinees usually take the DLPT5-Reading lower level and if they score 3 on the test, they become eligible to take the DLPT5-Reading upper level. The study was focused on the DLPT5-R lower level because all of the examinees will take DLPT5-R at the lower level first.

There are two types of DLPT5: the first one is the multiple-choice (MC) format, and the second one is the constructed-response test (CRT) format; both tests' formats are designed for all levels of the test. The DLPT5 MC (See Appendix A) response format is used when there is a large population who will take the language test, and the test will be scored by computer. The DLPT5 CRT (See Appendix B) is used when there is a small population who will take the language test, and it is scored by DLPT5-certified testers.

DLPT5 Reliability and Validity

The DLPT5 is a valid and reliable multiple-choice test and has been conducted and used at the DLIFLC for many years. "The integrity of the DLPT5 testing system

relies on test users' confidence in the tests. To ensure DLPT5 test validity and usability, standardized validation procedures are being put in place for ongoing evaluation of all current DLPT5 tests" DLPT5 guide (p. 79). Such validation procedures include review and analysis conducted by experts in testing and by experts in the ILR skill level descriptions. Furthermore, the DLPT5 is pre-tested through a large number of examinees and the data are analyzed and questions that are not functioning properly are removed.

Theoretical Foundation

The study investigates the relationship between the results obtained from a DLPT5-Reading test of Arabic and the results from a validated Can-Do-Scale (CDS) language self-assessment of Arabic reading ability. There are many theories that are directly linked to self-assessment and testing; these theories are Constructive Learning Theory, Multiple Intelligences Theory, and Social Cognitive Theory.

Constructivist Learning Theory

Woolfolk (2007) shared that the constructivist learning theory perspectives are grounded in the research of Vygotsky, Piaget, Gestalt, Dewey, Bruner, Bartlett and many other intellectuals. There is more than one constructivist learning theory, but most of them share two thoughts: the learner is active in constructing his own knowledge from previous knowledge or experiences, and social interactions are very important for the learner to construct new knowledge.

McMillan & Hearn (2008) pointed out that students need to use self-assessment when they are learning, to see what they know, and how much more effort they need to gain more knowledge, and to be successful. Students need to know when they are making mistakes and identify the learning strategies that work best for them. However,

classroom-based assessment conducted by teachers is still an important aspect in assessing student performance (Valencia, 2002). Accurate evaluation is important to identify what students have achieved and specify what additional work is needed to accomplish their learning goals.

Multiple Intelligences Theory

Ritchie (2009) states that Gardner has written about multiple intelligences and language aptitude, as learners who have a high level of ability will succeed better in learning a language. Armstrong (2009) cited that Gardner has mentioned in his book *Frames of Mind* (1983) the existence of at least seven intelligences; Gardner has also added an eighth intelligence and discussed the existence of a ninth intelligence. The multiple intelligences theory has been selected because improving students' ability in self-understanding, self-knowledge, and problem solving is at the core of self-assessment in this study.

Campbell, Campbell, and Dickinson (1999) cited Gardner's definition of human intelligence as "the ability to solve problems that one encounters in real life; the ability to generate new problems to solve; and the ability to make something or offer a service that is valued within one's culture" (p. xv). Campbell et al. (1999) stated that Gardner intended to create seven instruments to measure human intelligences at Project Spectrum as soon as he discovered the human intelligences. Then he realized that human intelligences existed in a vacuum. These intelligences are:

1. Linguistic intelligence: focuses on the mastery of language and on effectively using words in writing or orally.

2. Logical-Mathematical intelligence: involves using numbers and reason effectively, and in paying attention to logical patterns in calculating, hypothesizing, classifying, and categorizing things as examples.
3. Spatial intelligence: involves paying attention to perceiving mental images accurately and transforming and manipulating those perceptions in solving problems.
4. Bodily-kinesthetic intelligence: involve using the hands or whole body movement and physical skills in expressing feelings and ideas in balanced, coordinated methods.
5. Musical intelligence: involves ability in perceiving music, identifying, categorizing, converting, and expressing musical melodies, rhythms, and pitch.
6. Interpersonal intelligence is “the ability to perceive and make distinctions in moods, intentions, motivations, and feelings of other people” (Armstrong, 2009, p. 7).

Interacting with other people may include verbal and nonverbal communication as in facial expressions and speech, to respond effectively to certain interpersonal cues in practical ways.
7. Intrapersonal intelligence includes self-knowledge, self-awareness, self-understanding, self-discipline, and capability in adapting and acting knowledge, and accurately knowing one’s strengths, weakness, and motivations.

Campbell et al. (1999) referred to Intrapersonal intelligence as “the ability to construct an accurate perception of oneself and to use such knowledge in planning and directing one’s life” (p. 7).
8. Naturalist Intelligence (*nature smarts*) deals with the ability to recognize and classify various species, like plants, animals, and nature. It also pays attention to natural

phenomena (mountains, sunrise, and wind) and to environmental changes and inanimate objects and surroundings.

Social Cognitive Theory

Bandura, as cited by Woolfolk (2007), posits reciprocal determination, where the author believes that external and internal factors are important in social cognitive theory. Woolfolk (2007) also stated, “Environmental events, personal factors and behaviors are seen as interacting in the process of learning” (p. 330). The physical and social environment (the physical setting, feedback, consequences of actions, instruction, other people, models, resources), the personal factors (self-regulated progress, outcome expectations, attributions, attitude, progress self-evaluation, self-efficacy, expectations, beliefs, knowledge), and behavior (learning, verbal statements, motivation, choices, individual actions, goal progress,) are all influenced by and influence each other (Woolfolk , 2007).

McMillan and Hearn (2008) pointed out that self-assessment is an essential component of the constructivist and cognitive learning theories. Students will be able to self-monitor their learning and the ways they think, and they will be able construct their knowledge and meaning. All of these components are part of a self-assessment process where students need to organize, assess, and internalize their thinking in gaining knowledge. Students will need to use and connect the stored constructed knowledge with the new information, skills, and understanding. Self-assessment helps students’ to make connections between the information they have and themselves in meaningful ways; this process of connection encourages them to learn and increase their confidence and motivation, rather than making learning a mechanism of memorization and repetition.

Research Questions and Hypothesis

The following research questions and hypotheses guided the study:

RQ1: What is the correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

RQ2: What is the difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

RQ3: Is there a difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered?

H₁: There is a statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

Ho: There is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

Definition of Terms

Self-Assessment (SA): “the evaluation or judgment of ‘the worth’ of one’s performance and the identification of one’s strengths and weaknesses with a view to improving one’s learning outcomes” (Klenowski as cited by Ross, 2006, p. 1)

Defense Language Proficiency Test 5 (DLPT5): The fifth generation of the Defense Language Proficiency Test. (Elfiky, 2012)

Defense Language Institute Foreign Language Center (DLIFLC): Foreign Language Institute that provides resident instruction at the Presidio of Monterey.

Can Do Scale (CDS) for self-assessment: A self-assessment of foreign language comprehension proficiency test developed from the criteria of the Interagency Language Roundtable Language Skill Level Descriptions (ILR, 2012a, 2012b).

Interagency Language Roundtable (ILR) for Reading Skill Level Descriptions: a scale that is designed to measure language proficiency and describes different levels or degrees of Reading proficiency, ranging from 0 (No Proficiency) to 5 (Functionally Native Proficiency) (see Appendix C) (ILR, 2012a).

Interagency Language Roundtable (ILR) for Speaking Skill Level Descriptions: a scale that is designed to measure language proficiency and describes different levels or degrees of Speaking proficiency, ranging from 0 (No Proficiency) to 5 (Functionally Native Proficiency) (ILR, 2012b).

Language Proficiency Test: A test designed to measure a learner's ability to function in a language in real-life situations regardless of the type of education s/he may have had in the language (Elfiky, 2012).

Military linguist: A military person who is skilled in at least one foreign language in addition to his/her native language (Elfiky, 2012).

Limitations and Delimitations of the Study

Limitations

1. The results of the study may not be applicable to other institutes or universities, or other bilingual program because the subjects are military linguists.
2. The study may not apply to other language programs that are not using ILR skill level descriptions in evaluating students.
3. Participants may withdraw from the study because they cannot take the DLPT5-Reading test; students are military and may be asked to do other duties or even leave the program at any time. Participants' dropout can be addressed by increasing the target number of participants.
4. The CDS instrument used in the survey is written in English, so it might not be understood by students who do not speak English as their first language.

Delimitations

1. The students who will participate in the study are only in ages 18-40 years old.
2. The participants of the study are limited to military linguists who are studying the Arabic Basic Course at the DLIFLC in the three Middle East Schools, in Monterey, California.

3. The study is limited to students who are taking the DLPT5 Reading skill lower level (from 0-3 on ILR scale) of Arabic.
4. The study is limited to assessing the students' skill of Arabic reading comprehension.

CHAPTER TWO: REVIEW OF THE LITERATURE

Introduction

The purpose of the study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). To accomplish the study, the researcher will investigate the correlation between the two assessments: ratings obtained from the CDS (a self-assessment instrument) of Arabic reading ability and scores obtained from the valid and reliable DLPT5-R Test of Arabic reading ability. This chapter will discuss the relevant literature related to the study.

Theoretical Foundation

The Social Cognitive Learning Theory

There are many learning theories that are related to self-assessment and learning; the study is built on three learning theories, the Social Cognitive Learning Theory, the Constructive Learning Theory, and the Multiple Intelligences Learning Theory. Woolfolk (2007) motioned a definition of self-regulation “self-regulation as the process we use to activate and sustain our thoughts, behavior, and emotions in order to reach our goal” (p. 335). The ultimate goals of self-evaluation are getting the feedback from the teachers, construct, and build on what the students know and skills they have. To give the students the opportunity to be self-regulated learner, the student must know their ability, the knowledge they have, skill, the task, strategies for learning, the subject they want to learn about, the context where they will apply what they learned. Students’ self-evaluation will help them to self-regulate their learning. Woolfolk (2007) also discussed that “expert” students know their strength and weakness, how to cope with difficulties,

the best learning strategies that help them learn better, as they know their learning style, interests, and talents.

Woolfolk (2007) acknowledged that “involving students in generating evaluation criteria and evaluating their own work also reduces the anxiety that often accompanies assessment by giving students a sense of control over the outcome” (p. 342). Self-evaluation practices that maintain Self-Regulated learning do not threaten the learning process, because both are embedded in learning, encourage progress as the students enjoy participating in it and encourage them to seek challenging tasks because the students participation is low; self-evaluation process help the students find their mistakes and interpret them as an opportunity for learning (Woolfolk, 2007).

Woolfolk (2007) shared that one of the most important goals of teaching is preparing the student to be a lifelong learner; and in order to be a lifelong learner, students must self-regulate their learning and their education. Students must be motivated, knowledgeable, volition for learning that grants the skills they need to learn effectively and independently. Knowledge includes the students’ understanding of themselves, the learning tasks, the subject contexts, and learning strategies they need in mastering new skills. Woolfolk (2007) also noted, “Motivation to learn provides the commitment and volition is the follow through that combats distraction and protects persistence” (p. 366).

As cited in Woolfolk (2007), Zimmerman provided three phases of self-regulated learning which is described as the learning cycle of self-regulation “forethought (which includes setting goals, making plans, self-efficacy, and motivation); performance (which involves self-control and self-monitoring); and reflection (which includes self-evaluation

and adaptations, leading to the forethoughts/planning phase again)” (p. 366). Grabe (2009) states that

the social-cognitive theory combines the influence of cognitive abilities, environmental factors and behaviors in a given situation. Self-perception centers on the concept of self-efficacy, a person’s belief about his or her ability to learn or perform actions successfully. Self-efficacy is a major component of social-cognitive theories. How people feel about themselves and their abilities to learn or perform affects effort expenditure, persistence, and learning. (p. 178)

Grabe (2009) also found that,

self-efficacy is important in predicting learning, motivation and achievement. Students can improve self-efficacy by setting immediate and more limited goals, evaluating their learning progress; receiving useful feedback, learning to connect success with effort and ability; and learning to self-monitor and check progress. Self-efficacy is also related to self-regulation as an expected outcome of strong self-efficacy in academic context. (p. 178)

The Constructive Learning Theory

Woolfolk (2007) discusses psychological constructivism that focuses on how the learner uses the resources, information, or assistance from others, experiences and problem-solving strategies, and improving their mental ability in constructing new knowledge. The social constructivism intellectuals look at learning as a way to increase our abilities to communicate, and participate with others in meaningful cultural activities.

Woolfolk (2007) stated, “Vygotsky believed that social interaction, cultural tools, and activities shape individual development and learning” (p. 346). Woolfolk (2007) also mentioned that Piaget thinks that constructing knowledge comes from the internal direction, in organizing, categorizing, transforming, and recognizing previous information or experience, and he believes that discovering and exploring of knowledge are more important than teaching. Students’ self-evaluation is a strategy of discovery and exploration of knowledge and skills they have. The students will be able to build on the

information and feedback they get to construct new knowledge and experience in learning foreign languages.

Traditional and standardized assessment is frequently criticized because it focuses on memorization and recalling what people learn. When we look at assessment as it deals with vital aspects of learning, it will have great and constructive positive influence in improving instruction. This how students self-assessment provide more information about their understanding and comprehension (Campbell, 1999).

Multiple Intelligences Learning Theory

Gardner (1983) stated that humans have unique intelligences that vary from one person to another. Human species are different culturally and naturally, from each other; these intelligences are not limited to the intelligences. Gardner has identified and categorized humans' intelligences. Gardner also criticizes I.Q. tests as they do not measure completely human talents; most I.Q. tests have limited measures to test human abilities. Gardner also has noted that every human intelligence include a number of sub-intelligences; as a musician who can sign, read, conduct, criticize and play music. Gardner (1993) stated that multiple intelligences can be subdivided and rearranged in a certain way. People may have different intelligence profiles that they are born with. Human intelligences work collaboratively and connectedly in solving problems.

To achieve the goals of having, autonomous, lifelong learners, students must have the opportunities to be active self-assessors, self-monitor their achievements, manage what they learn, and in critiquing their weakness and strengths, and to recognize the process of how they learned, and what they need to learn to achieve their goals. Writing journals, portfolios, peer assessment, informal student/teacher dialogue, self-reflection

sheet are good examples of active assessment of the self. Assessment becomes more meaningful and more relevant when students perceive and start reflecting internally about themselves. Students can construct their personal understanding and perception of the subjects they learn (Campbell et al., 1999). Campbell et al. (1999) stated that using the appropriate intelligences in assessing the students' ability would minimize the threat that can be caused by traditional assessment and traditional testing.

Current Literature Related To Study

Baniabdelrahman (2010) conducted a research study on High School students who were learning English in Jordan, and found that self-assessment affects reading performance positively. Wan-a-rom (2008) conducted a study on high school students learning English in Thailand and found that self-assessment (SA) directs learners to an appropriate reading level for extensive reading curriculum. LeBlanc and Painchaud (1985) conducted a study on first-year students of English and French at a University of Canada, and found positive correlations between SA and standardized proficiency tests of the reading and the listening skill. There have been no studies investigating the relationship between Arabic reading ability self-assessment and the DLPT5-R skill of Arabic language. Elfiky (2012) conducted a correlation study between Oral Proficiency Interview (OPI) of speaking skill and students' self-assessment. His study revealed a significant correlation between CDS and OPI ($r = .272, p < .05$). In addition to that, the percentage of perfect agreement between CDS and OPI was 58%.

Brantmeier, Vanderplank, and Strube (2012) revealed in their study that, as measured through multiple choice reading items, self-assessment of reading ability as a second language is not a correct predictor of the following reading performance when it

is measured with students self-rating scales. The study has also revealed that students can give an estimate of their reading ability accurately. The study showed that self-assessment instrument could benefit the students' success in their study as they get involved in the process of learning and assessment. Self-assessment may also document the individual students' performance and their learning development over time. The study has also revealed that advanced students know when they are weak or strong at different language skills.

Self-Evaluation

Woolfolk (2007) mentioned that self-evaluation is more difficult than self-recording because self-evaluation involves judgment of the quality of skills that the student has about himself/herself. Students can judge and evaluate their behavior if they know and learn the standards in evaluating and judging their performance in such skill or a product. Woolfolk (2007) stated, "Self-evaluation can accompany self-correction. Students first evaluate, then alter and improve their work, and finally, compare the improvement to the standards again" (p. 235).

The Importance of Self-Assessment

Chen (2008) conducted a comparison study between the students' self-assessment and teachers' assessment of English oral skill performance as a second language. The study showed that feedback, training, and practicing self-assessment increased the accuracy of students' self-assessment, and it showed that students achieved their learning objectives and goals. The first phase of the study results showed that students' self-assessment differed significantly from the teachers' assessment, but in the second phase

of the study showed that students' self-assessment and teachers' assessment were closely aligned.

Blanche (1988) noted that in eight studies conducted by Lee and Low, 1981, 1982; Fok, 1981; Bournemouth Eurocentre, 1982; von Elek, 1981, 1982; Heindler, 1980; and Heidt, 1979, "it was found that applying self-assessment have increased students' motivation" (p. 82). Blanche (1988) noted that other studies have shown largely that students who overestimate are usually the weaker ones than high achiever students.

Jiang (1999) found a positive correlation between the students' first language self-assessment for reading and self-assessment for second language vocabulary skills at community college. Jiang (1999) cited in his research Blanche (1988), who stated that self-evaluation is important in increasing students' motivation for learning. In implementing self-evaluation, students become more aware of their responsibilities, own difficulties and their own progress, and minimizing the teachers' role and dependence in giving opinion totally; students give hints to the teachers about their individual needs when they use self-evaluation (Shen, 2002).

Blanche (1988) also revealed the importance of self-evaluation in the early stages of studying a new language, curriculum used, and how practicing self-evaluation is important and help in promoting language acquisition in classroom. Linguistic skills and the materials educators involve in evaluating students plays an important role of the self-evaluation accuracy. They also noted that standardized tests are important as self-assessment tool. There are many things involved in self-assessment that make it complicated and which incorporated the students' varied cultural backgrounds, the difference between the teachers' values and the students' values; the absence of having

valid criteria that teachers and students could use in making better decisions make self-assessment complicated.

Jiang (1999) noted that self-judgment could be a consistent assessment tool that teachers depend to evaluate students who learn English as a Second Language (ESL) if students have sufficient guidelines in knowing the required expectation from them, and required proficiency level they need to achieve.

Self-assessment is important because students will be able to identify their learning strategies and increase their understanding and achievement (McMillan & Hearn, 2008). Language Self-Assessment questionnaires are good opportunities to understand what students know and their ability to learn a language (LeBlanc & Painchaud, 1985). Using self-assessments improve students' control over achievement of their learning goals, encourage, and promote students' self-awareness to learn (Butler & Jiyeon, 2006). Self-measure is a good opportunity to know more about students' progress and get feedback, then teachers and students have time to implement changes in learning and teaching approaches before they get their final tests (Shohamy, 1992). Self-assessment will increase the students-teachers interactions and make communication better in opening up direct communication lines (Byers, 2010).

Language Use in Self-Assessment (Methodology)

Blanche (1988) suggested that the questionnaires used with learners self-assessment should be in their native language and not in the language that they are learning, except when their linguistic development is in a fairly advanced phase.

Students Self-Placement

Royer and Gillis (1998) noted that because the placement tests that the students should take are unreliable, students should self-evaluate their abilities. It is important to know more about students from more than a single test result; educators may ask the students about self-image (e.g. describing their strengths, weaknesses, and changes they have in their study style). Students' self-placement and evaluation are important to students because they are responsible about their own education.

Royer and Gillis (1998) also noted that, at the college level, placement tests failed to recognize the students' remedial lessons before they take the standard composition classes and spread frustration between teachers and students. Students may be placed in courses that are not counted credit in college. Teacher had to deal with students' frustration by giving them a replicate of the placement essay in the beginning of the course and moving those students to appropriate classes despite of their will. Since the test results did not succeed in increasing the students' self-confidence. They believe that self-placement system in their school is working but it is still early to represent precise conclusions regarding it.

Royer and Gillis (1998) conducted a study in 1996 where they noticed that 59% of the students reflect self-image and self-placement and 27% of the students reflected judgments from outside. The study showed that the majority of students selected EGG 098 class is because of their own self-view and judgment.

Perceptions of Self-Assessment

Ross (2006) found that despite the negative view of how teachers look at self-assessment as being accurate and valuable, research found the opposite of this negative

look at SA. Ross (2006) conducted research on the reliability, validity, and utility of self-assessment and found the opposites of these claims “self-assessment produces consistent results across items, tasks, and short time periods” (p. 1). He also found the SA importance in improving the students’ achievement and behavior. “Self-assessment provides information about student achievement that corresponds only in part to the information generated by teacher assessments” (Ross, 2006, p. 1). Teacher’s training students how to assess themselves can enhance and strengthen self-assessments.

Self-Assessment Requirements

Ross (2006) assured that, we need to meet three conditions before getting the whole advantages of self-assessments: First, having an open discussion between students and teachers about self-assessment criteria; Second, dialogue between students and teachers should focus on proofs for assessments; Third, teachers-students collaboration or only by students themselves in self-assessment adds a grade.

According to Wiggins (1993) Students should assess themselves in every major assignment they take. Self-assessment is considered a valid form of assessment; students develop higher order thinking skills, self-assessment uses transparent and clear criteria that are available for everybody, students have time for feedback, modification and progress (as cited in Ross, 2006). Blanche (1988) noted that students will be able to give accurate answers about the language competence when we use items that contain linguistic situation used in surveys (as cited in Byers, 2010).

Distinction between Self-Assessment and Self-Evaluation

Gregory, Cameron, and Davies, (2000) noted that some teachers think that the distinction between self-assessment (informal judgment about achievement) and self-

evaluation (judgment for grading) is helpful (as cited in Ross, 2006). On the other hand, according to McMillan (2004) some teachers think the distinction between both is not helpful and they use both terms interchangeably (as cited in Ross, 2006).

Reasons to Use Self-Assessments

Ross (2006) stated that teachers use self-assessments for many reasons; students' involvement in assessing their work, the criteria, and standards used will increase their engagement, awareness and their interest in learning. Self-assessment provides important information that is difficult to find through other kinds of assessments. For example, teachers can ask students on the time they spend in doing a task. Some teachers think that using self-assessment is much cheaper than other methods.

Self-Assessment Definition

The definition of self-assessment according to Klenowskis (1995) is “the evaluation or judgment of ‘the worth’ of one’s performance and the identification of one’s strengths and weaknesses with a view to improving one’s learning outcomes” (as cited in Ross, 2006, p. 1). Blanche (1988) mentioned that the first appearance of the term self-assessment in literature has appeared in the year 1976 and was referred to as by “self-rating”, “self-appraisal”, “self-control”, etc... Self-assessment is related to learners’ autonomy where students will not depend entirely on the teachers in conducting evaluation. Students will self-assess themselves accurately and at the same time students will make their teachers mindful of their different needs. It is very important for students to know what progress that they have achieved from their studies, and the skills that they need to master and acquire. Without this knowledge, it will be difficult for students to learn efficiently.

Heilenman (1990) mentioned that Upshur (1975) was one of the beginners who studied and provided a rationale of using self-assessment in measuring second language acquisition. “Learners have access to the entire gamut of their success and failures in the use of the second language, whereas any test of actual language use, of practical necessity, can sample only a small portion of that ability” (p. 174). Self-assessment is not used for certification. Many learners use it for learning informally and in getting information about how they learn (Dickinson, 1987).

Self-Assessment Methods

One of the methods in self-assessment is looking at students’ strengths, weaknesses, opportunities, and threats (SWOT) they have or feel in learning a language. In self-assessments, teachers can also look at political, economic, social, technical and legal forces (PESTL) that are involved in a student’s learning (Bannock, Davies, Trot & Uncles, 2003).

Assessment Categorization

Assessment has been categorized in terms of “(a) norm-reference and criterion-reference testing, (b) formative and summative assessment, (c) formal and informal assessment, (d) proficiency, achievement, placement, and diagnostic assessment” (Grabe, 2009, p. 353).

Need for Self-Assessment

Royer and Gillis (1998) stated that some teachers are not comfortable with traditional methods in assessing students “our discomfort with traditional placement methods arose from an uneasy feeling of impropriety” (p. 63). Within an hour or two, teachers and administrators are about to make a big decision for many students, and they

did not know about the courses that students should take; but after two hours everybody knows and no matter how careful and how accurate who made these decisions are but the decisions were hasty and excessively quick (Royer & Gillis, 1998).

Self-Assessment Reliability and Validity

Although the benefit of self-report or self-assessment is a good alternative method in exploring and judging the students capabilities in performing the new language they are learning, there are still some concerns and debate about the precision and the validity of this measuring method when there is a need to have important decisions (Byers, 2010). “Self-estimation on foreign language proficiency has proven to be a reliable measurement of language proficiency” (Beerkins, 2010, p. 86).

Byers (2010) stated that, the University of Tennessee at Chattanooga has conducted a self-report survey about the student foreign language acquisition of Reading, Listening, and Speaking skill, and cultural understanding, and writing. The purpose of this survey was to meet the Foreign Languages and Literatures Department requirements, and to capture students’ progress in a different way of measuring. The survey was 35 “I Can” items that measure the students understanding and using foreign language has helped the instructors to determine if their students are progressing and meeting the language requirements.

According to LeBlanc and Painchaud (1985), self- report method can take place on different locations and on different times and not be limited to class setting, which make the students feel more comfortable and not to be afraid of standardized testing consequences, the students do not need to cheat on tests because self-report results will

not be graded and it will not affect them (as cited in Byers, 2010). Byers (2010)

elaborates on the use of self-assessment in Foreign Language Departments,

Although there are various benefits when using self-assessments, there are some reliability and validity issues connected with educational assessments due to the various differences in students' performances and abilities. However, by using additional methods, Foreign Language Departments can create tools that can better judge if students are able to meet the goals established in their outcome statements. (p. 2)

According to Brindley (2001), "Not only do assessments of language performance need to meet the requirements of validity and reliability, they also need to be practically feasible" (p. 2).

Byers (2010) noted that the reliability of test or assessment is concerned with what degree the measurement error effects tests and the scores, i.e. different scores can be affected by factors not linked to the ability being evaluated e.g.(students are guessing, fatigue, got instructions before assessment, etc.). These factors may create inconsistent results by students. To establish consistence stable results over a period, different methods can be implemented. Consistence and reliable results over a period of time can be assessed in terms of test-retest study; where the same assessment can be given to one group but with two different aspects or giving the same assessment but with two equivalent assessments forms (Byers, 2010).

Blanche (1988) stated that some researchers revealed those self-assessment results "may often be affected by subjective errors due to past academic record, career aspirations, peer-group or parental expectations, lack of training in self-study and self-management, etc...." (p. 81). It is important for the students' learning development to practice self-directed learning and self-directed assessment, both of which are considered prerequisites of learning development process (Blanche, 1988).

Steps to Assure Reliability of Self-Assessment

DeVellis (2003) noted that, researchers need to create an estimate of consistency to make the self-assessment instrument reliable; this could be done in creating a reliable scale, which refers to the variance between students' exact grade and scores obtained from self-assessment (as cited in Byers, 2010)

Byers (2010) noted that self-assessment results are likely subjective, it creates problems to the reliability of using this method; reliability of any measurement depends on condition, circumstances, and the assessment purpose. Most standardized testing, such as multiple choice exams can obtain high reliability when they are compared to self-assessment and portfolios because multiple choice tests has one correct answer or right or wrong answer. Moss (1994) assured that, to obtain high reliability in assessment, there should be more structured, and controlled performance assessments that measure students' performance in a generalized matter (as cited in Byers, 2010).

To obtain reliable results in research, there should be a shorter phase between assessments and measurements, and more balanced assessment items and tasks. Research shows that, to obtain greater and higher consistency in the students' responses in skills evaluation, students should be trained on how to self-assess themselves, have clear instructions, and get brief directions before taking the self-assessment (Ross, 2006). Blanche (1988) noted that, obtaining accurate and consistent results are very important for the students who want to excel their learning and understanding and for those who want to let teachers know about their strengths and weaknesses; self-assessment assures that the students are active in evaluating themselves and it is not only based on the teachers judgment and views (as cited in Byers, 2010).

Literature Research Review

Woo (1995) found in a study that students self-assessment of language proficiency prior language experience is the best predictor of the DLPT III formal results in learning Korean language at the DLIFLC for reading and listening skill. Pinto (2009) found in her qualitative and quantitative study titled “A Study of the Seventh Grade Students’ Reading Comprehension and Motivation After Explicit Instruction in Self-Assessment and Metacognitive Reading Strategies”, those students who were given detailed and clear instructions in self-assessment and metacognitive reading strategies, established a capability to generalize educated metacognitive reading strategies to the different script. The study (Pinto, 2009) also revealed that students, who established metacognitive reading strategies, learned self-assessment, encouraged active, self-regulating learning and thinking, increased their motivation, and reading comprehension. Wolochuk (2009) found in her study “that a significant correlation ranging from moderate to low with the three self-assessment variables (understanding spoken English, writing, and reading)” (p. 53). The study also found a “positive correlation between self-assessment of reading skill and Test of English as a Second Language (TOEFL) results of reading skill” (Wolochuk, 2009, p. 53).

Yuko and Lee (2010) revealed in their study that there is a minimal positive effect of self-assessment and self-confidence on 6th grade students who were learning English as a foreign language in South Korea. The study also showed that teachers and students look at self-assessment effectiveness differently and this basically depends on how they look at context of teaching and learning, and on how each teacher viewed assessment.

Butler and Jiyeon (2006) examined in a study the validity of Korean students' self-assessments of their oral performance in English in a Foreign Language at the Elementary School (FLES) level,

the results indicate that if self-assessments are administered in an on-task format, students can self-assess their oral performance more accurately than they can in an off-task format. It was also found that the on-task self-assessment was generally less influenced by student attitude/personality factors than was the off-task self-assessment. (p. 1)

Self-Assessment Process

McMillan and Hearn (2008) assured that self-assessment increased students' motivation and engagement in learning and made learning more meaningful. Self-assessment has an influential impact on the students' classroom performance, their accountability and in guiding their education.

The process of self-assessment combines three components in a cyclical constant manner:

1. Students awareness of their education, thinking or actions (self-monitoring),
2. Students know how to judge themselves toward reaching the learning goals and targets (self-Judgment).
3. Students identification of their learning strategies and needs; and then apply these techniques in correcting and improving their performance.

Self-Assessment Implications for Practice

Students' self-assessment in classroom improves their awareness of which metacognitive approaches are better to use and when to apply them. In order for the students to learn these skills and evaluate their work, teachers need to have clear lessons, learning objectives, goals, and evaluation criteria. As a result, students will be engaged,

actively involved, and connected in the learning process and learning outcome. Teachers are responsible in involving students in learning, and passing and shifting evaluation to students through scaffolding; scaffolding require teachers to maintain high expectations from students self-evaluation; teachers should work as trainers and advisors as students are learning from their personal knowledge and experiences (Joyce, Weil, & Calhoun, 2005, as cited in McMillan & Hearn, 2008).

Self-Assessment Studies

Sternberg (2002) defines Self-monitoring refers to the students when they keep monitoring and tracking their own progress. Sternberg (2002) cited that Morgan (1985) found in his study that students who self-monitored sub-goals, and kept tracking of their progress in every step to complete the duties were more successful and scored better in the tests than students who did not self-monitor their time on studies or have distal goals; the study has also revealed that students who self-monitored time spent more time in studying than the other group. The sub-goal self-monitoring students have showed more intrinsic interest in the course than the other group.

Sternberg (2002) noted that students' goals are considered one that motivates students; there are many things that contribute in motivating students as the students' needs. Abraham Maslow (1970) mentioned that students need to gain achievement, power and affiliation fits in the theory of motivation (as cited in Sternberg, 2002).

Motivational Theory (Maslow's Hierarchy of Needs)

Sternberg (2002) mentioned that Maslow has argued seven needs before humans can reach self-actualization; self-satisfaction, self-monitoring, self-evaluation, and seeking personal growth. These needs starts when the physiological needs are satisfied,

then safety needs , belonging and love needs, then self-esteem needs, then need to know and understand, then aesthetic needs and final the self-actualization needs. To reach the self-actualization, we need to become aware of the inner self and inner feelings.

Cooley (1982) clarified self-esteem “Self-esteem refers to the value a person places on himself or herself. Self-esteem is related to self-concept or one’s ideas about one’s attributes and abilities” (as cited in Sternberg, 2002, p. 373). Our own judgments of ourselves are not the only thing that affects self-esteem, but also when others judge and evaluate us.

Reading Assessment

Grabe (2009) noted that reading assessment is a great tool that informs administrators, teachers, researchers, policy makers about students. Reading assessment could create a significant power that benefit the education atmosphere or it could harm it harshly. We should be careful, mindful, and pay more attention, when dealing with reading assessment and its consequences.

Grabe (2009) stated that reading assessment gives feedback of the students’ knowledge, skills, and procedures that students use to obtain the reading ability. Reading assessment can be categorized in many different ways and it is based on different theories. Generally, learning theories have been categorized in many ways and have their unique purposes and frameworks (p. 353):

1. Norm-reference and criterion reference testing
2. Formative and summative assessment
3. Formal and informal assessment
4. Proficiency, achievement, placement, and diagnostic assessment.

Grabe (2009) also proposed five purposes of reading assessment (p. 353):

1. Reading Proficiency Assessment (Standardized testing)
2. Assessment of Classroom learning
3. Assessment of learning (supporting student learning is the purpose)
4. Assessment of curricular effectiveness and
5. Assessment of research purpose

Informal Assessment

Grabe (2009) stated that self-reporting measures are considered one of the wide informal assessments choices that teachers use beside observation, portfolios, students' progress chart, etc. Teachers use multiple methods or options of informal assessment to obtain a clear picture and monitor the students' progress and their reading ability. Grabe (2009) noted that informal assessment should be more objective and more personalized; to make the informal assessment more objective and fairer to students, teachers need to know that these kinds of assessment grades will be used for job advancement, promotions, or placement.

Grabe (2009) further noted that self-assessment is considered an important element of informal assessment. Self-assessment requires from students many things; (a) Students are required to acquire important information about their learning, map and monitor their own progress and their advancement, (b) know what they read and reason behind their readings, (c) justify their reading options, choices, their objectives and goals, (d) create a reading strategies list that they use now or would like to use, (c) evaluate their own reading portfolios. He stressed that self-assessment is very important in creating better self-awareness and promoting continual learning when students' self-assessments

for reading is discussed, reviewed and evaluated. Self-assessment creates a motivating atmosphere for students to expand, explore valuable learning strategies, and develop (Grabe, 2009). Black and William (1998, 2005) and William (2007/2008) mentioned that more than 4,000 studies conducted over the past 40 years have shown the use of assessments doubles the rate of students' learning (as cited in Grabe, 2009).

Informal assessment is an important and great opportunity that allows teachers to engage many students in feedback. To allow self-assessment for learning, students will indicate hints and signs of the difficulties they have, so their teachers can help solve these difficulties. Teachers responses based on assessment should address the skills that are needed for students learning to improve and students to achieve. Teachers should also promote students to be aware of what successful results look like and grant the opportunities for it (Grabe, 2009).

Training in Using Self-Assessment

To promote and encourage students for positive learning practices, teachers should not use informal assessment as a main source and basis for grades and evaluation of the students. Assessment provides direct data for teachers to adjust their instruction style that is based on the students' learning needs and interests (Grabe, 2009). Grabe (2009) stated that it is not difficult to train teachers as experts to use informal assessment with their students appropriately, and can be used to support students' learning continually. Teachers need to investigate and know more about the assessment standards used with students. Discovering these standards could be through study group discussion and constructive feedback to know the technical terms specifications, and consequences of using assessment.

Teachers can involve themselves in active research projects which, provide information and examine assessment for learning purposes, and how it impacts the students. Students can also be involved in training themselves on how to utilize self-assessment for reading comprehension, address their learning difficulties, and increase their performance (Grabe, 2009).

Self-Assessment Verses Other Tests

Dickinson's (1987) findings show there are some satisfactory indicators where educators can self-assess themselves accurately. Dickinson (1987) cited Oskarsson's 1984 research where he found an encouraging moderately consistent agreement between external criteria and self-assessment in language learning. Self-assessment can be used for many reasons to include self-monitoring, diagnostic testing, and placement testing. Self-assessment is very important in making self-autonomy learners and in making accurate and appropriate judgment about their performance. Self-assessment focuses on emphasizing learning and the process of practicing knowledge rather than focusing on the results and the products (Dickinson, 1987).

Justifications for Using Self-Assessment

Even though there are research studies that support self-assessment as accurate and acceptable, some may not be convinced enough to implement it. Some teachers and specialists may feel they can give accurate assessment more often than the students themselves (Dickinson, 1987).

Dickinson (1987) noted the reasons behind using self-assessment in learning are:

1. First, assessment that leads to evaluation is one of the learning objectives where students' training is very important for learning, and students become more effective autonomous learners, independent and self-monitoring learners.
2. Second, self-assessment is considered an important element that leads to self-direction, and learner-centered. Students will be responsible for their own education and be involved in making decisions.
3. Self-assessment helps teachers in reducing, alleviating, and lessening the assessment burden on them. For example, teachers do not need to give counseling to students and which can be concluded and conducted easily by the students themselves.
4. Students' self-assessment is very important because teachers may not be available to test students all the time. Students may also be studying and learning on unpredicted time and on varied things and where they need a kind of assessment that present feedback information.

Dickinson (1987) stated that in order for the students to monitor their progress, see their strength and weaknesses over a period of time, it is important to keep a record of all self-assessments results. Dickinson (1987) noted that self-assessment, studying materials, and tests should be related to obtain validity "If learner-constructed tests are closely related to the learning material used-the course book, for example- then the content validity may be protected to some extent" (p. 149).

Summary

This chapter has reviewed the learning theories related to self-assessment in education; these theories are the Social Cognitive Learning Theory, the Constructivism Learning Theory, and the Multiple Intelligences Learning Theory. These theories will act

as the foundation for this study. This chapter discussed in detail the reasons behind selecting these learning theories. It also discussed the current literature related to this study, the importance of self-assessment, the negative look of self-assessment, the validity, and reliability of self-assessment, language use in self-assessment, age, gender, personality, and social class in self-assessment. This chapter also reviewed the formal and the informal assessments, and training in using self-assessment. Chapter Three will discuss the research methodology and design of the study. It will also discuss the selection of participants of the study, their selection, and their numbers. Chapter Three will include the instrument used in this study, the validation of the instrument, data collection and data analysis and ethical assurance on the study.

CHAPTER THREE: METHODOLOGY

Introduction

The purpose of the study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). Through this study, the Can-Do-Scale (CDS) Reading self-assessment instrument was validated by an expert panel and then it was tested in a pilot test and a full study, against the reliable Arabic DLPT5-R test. Participants of this study are Defense Department military students who are native speakers of English and who are learning Arabic as a second language.

To conduct this study, the researcher investigated the correlation between the two assessments: ratings obtained from the CDS (a self-assessment instrument of Arabic reading ability, adapted from the ILR website) and scores obtained from a valid and reliable Arabic DLPT5-R test. This chapter provides the detailed methodological outline of how the study was executed. The research design and approach will be discussed followed by the setting, and description of the sample. The data collection and operationalization of the variables will be presented along with the data analysis, instrumentation, validity, and reliability. Protection of human subjects will be discussed and a summary will conclude the chapter.

Research Method and Design

For this particular research study, a quantitative correlational research study was found to be appropriate. Other methods that were considered included the experimentation research design, and mixed methods approaches. The qualitative research design is most appropriate when the purpose of the study is to investigate the in-

depth experiences of the participants or to analyze qualitative data such as interviews or open ended survey response questions (Marshall & Rossman, 2008). Mixed methods research would be appropriate if the goal of the study were to explore the phenomenon through qualitative techniques such as interviews about lived experiences and substantiate it with quantitative data through survey and questionnaires.

The purpose of the study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). To accomplish the study, the researcher investigated the correlation between the Arabic reading proficiency test (DLPT5-R) and a validated Can-Do-Scale (CDS) language self-assessment instrument of Arabic reading ability that was developed and adapted from the ILR website; the study determined the degree of validity and reliability that exists for the self-assessment instrument as measured against the DLPT5-R Arabic Test.

Since this study's goals and purpose aligned with the correlational research design, the correlational research design was the best choice (Babbie, 2012). The quantitative research design is most appropriate when the variables under investigation are numerical in nature (Balnaves & Caputi, 2001). The variables in this study were ordinal as a result of scoring procedures on the survey that will be discussed in the instrumentation section. A non-parametric test such as the Spearman's rho correlation analysis was used to calculate the correlation between the independent and the dependent measures because the CDS and DLPT5 scores represent ordinal data rather than interval data (e.g., 0, 0+, 1, 1+, 2, 2+).

Significance normally refers to the statistical determination that the test statistics for the variables in the correlation analysis meet a minimum requirement to be deemed non-zero, one out of 20 times (Babbie, 2012). For this assessment, the determination of statistical significance was not sufficient since validity, reliability and strength of the correlation was tested. Finally, as there are no qualitatively collected variables of interest for this investigation, the mixed methods approach was also not appropriate. Therefore, the quantitative correlational approach was the most appropriate method for conducting the validity testing of the CDS self-assessment of Arabic reading ability and the students' scores on the Arabic DLPT5-R test. As indicated earlier, three research questions explored in the study were:

RQ1: What is the correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

RQ2: What is the difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

RQ3: Is there a difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered?

H₁: There is a statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

H₀: There is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

Participants

The participants of the study included 153 U.S. male and female military students from the four branches of service: Army, Navy, Air force, and Marines in the three Arabic Middle East Schools. The participants engaged in studying Arabic language for 63 weeks at the Defense Language Institute Foreign Language Center (DLIFLC) in Monterey, California. The study included participants between 18-40 years of age. Students learned in the same classrooms and attend the same Arabic language program. Students varied in work experience, educational degrees, educational experiences, and experiences in language learning. The educational background varies with some students with a high school diploma, others with associate degrees, some with graduate degrees, or some with post-graduate degrees.

The students participated in an Arabic language learning program where no more than six students were present in any one class. Students received between six to seven hours of instruction per day. Students may have also received extra instructional hours and special-assistance instructional hours based on their needs from 6:30-8:30 pm, to improve their language skills. The basic instructional program consisted of teaching students reading, listening, speaking, transcribing, and writing, translating from Arabic to English, and translating from English to Arabic. One hour per week, each student had the chance to teach other students Arabic in the classroom. This is called the Leaders In Front Teaching (LIFT) program.

This study used the most commonly applied form of non-probability sampling: convenience sampling. This form of sampling was employed since the classes were given regularly and all students could easily be asked to participate in the study. The target number of participants for this study was 153. However, Keuhl (2000) purported that for quantitative studies, a minimum recommended power can be targeted at 80% which reduces the total sample to 99 total students. Moreover, a power of 80% ensures that the statistical analyses could provide valid conclusions with regards to the total population (Moore & McCabe, 2006). The final sample was derived from those participants who voluntarily completed and returned signed consent forms, which were distributed before collecting any data from the students, and participated in the study.

Operationalization of the Variables

This study consisted of two key variables and six control variables. The two key variables were the Arabic DLPT5-R test scores (criterion variable) and the Arabic Reading CDS scores (predictor variable). The control variables were: the highest

education level completed, military branch, military rank, gender, age, and previous experience with language learning. It must be noted that for this study, that clustering in hierarchical fashion teaching teams within classes and school was not considered.

Can-Do-Scale Score. The CDS score was the independent variable and was a six-level ordinal variable. The six levels correspond to final grades on the CDS test: 0+, 1, 1+, 2, 2+, and 3.

DLPT5-R Score. The Arabic DLPT5-Reading score was the dependent variable that was predicted by the CDS score given the control variables.

Military Branch. Military branch was a categorical control variable.

Level of Education. The highest level of education was an ordinal variable with the following levels: non-high school graduate, high school graduate or GED, associate's or technical degree (two-year College), Bachelor's degree (four-year College), or Master's degree.

Previous Experience in Language Learning. The previous experience of the participants was a dichotomous control variable indicating whether or not the participant had been in a language learning class or in another country where they learned another language before.

Instrumentation

This study included the use of two instruments. The first was the validated CDS self-assessment instrument of Arabic reading ability, which was adapted from a self-assessment test of foreign language reading proficiency available on the ILR website. The second was the Arabic DLPT5-R final exam currently used to measure third semester students' reading proficiency levels. The original instrument from which the

CDS was developed was based on the Interagency Language Roundtable Language Skill Level Descriptions-Reading (ILR) criteria (see Appendix D). The ILR criteria and the original version of the self-assessment are in the public domain. Elfiky (2012) noted that the study self-assessment instrument source did not require prior permission to use it. The original version of the CDS (the reading self-assessment instrument available at the ILR website) did not undergo validation or reliability testing. As a result, a validation process using triangulation technique was conducted to ensure that the native speakers of English who will participate in the study will be able to understand the instructions. As defined by Yin (2013), triangulation is the technique used to ensure the convergence of data collected through using multiple sources.

In this study, the triangulation technique was used to ensure the validity of the CDS self-assessment instrument of Arabic reading ability through gathering the results of existing studies which used the same survey instrument. The triangulation technique is used to achieve consistency in findings from different sources. Therefore, findings from existing studies were compared to the data collected in this study to ensure that the participants were able to understand and answer the items in the survey instrument. Burns (1999) noted that a students' language level is of critical importance when they answer questions related to tests, surveys, and interviews.

The original self-assessment instrument for reading proficiency at the ILR website consisted of 21 “can-do” statements that covered five levels of proficiency from level 0+ to 4. The original instrument was developed based on a dichotomous response metric of either yes or no. If the participant thought the statement described their ability incompletely, the answer would be no. For the current study, the survey responses were

expanded to a five-point Likert scale to give the students the opportunity to more accurately self-assess their reading ability as follows: (1) quite easily; (2) easily; (3) with some difficulty; (4) with great difficulty; (5) not at all.

The original version of the reading self-assessment instrument available on the ILR website does not include plus-levels items; and because the Arabic DLPT5-Reading test does include plus-level questions, the Arabic reading self-assessment instrument of this study was expanded to 42 CDS items that included plus-level items (see Appendix E). These newly added plus-levels items were taken from the ILR skill level descriptions of reading ability and from the DLPT5 guide. Permission was requested to add these plus-levels items to the new reading self-assessment instrument.

The survey consisted of two sections: the first section of the survey contained questions about the participant's demographics like name, gender, age, rank, language experience, military service, educational backgrounds and degrees. The second section of the survey consisted of 42 CDS items (Arabic reading self-assessment) that represent six levels of Arabic reading proficiency; these CDS items were constructed with a bottom-up approach from level (0+ to 3).

Validity and Reliability

The validation procedure for the CDS (Arabic reading self-assessment instrument) went through several phases. The study was e-mailed to the Evaluation and Standardization Division panel who were experts in the ILR language skill level descriptions, to inspect, review the reading self-assessment instrument, make ILR terminology simple to understand, review grammar, and lastly validate. First, the researcher talked with the validation panel to clarify the study and what was expected

from the members of the panel to do, and the steps in the validation procedure. After the discussion with the validation panel about the study, the original Arabic reading CDS (self-assessment instrument) was sent to the panel by e-mail.

Second, the validation panel recommended changing and expanding the self-assessment response scale from yes or no to a five-point scale. For example, for CDS self-assessment statements, participants select among five alternative options of the reading self-assessments: (1) quite easily; (2) easily; (3) with some difficulty; (4) with great difficulty; (5) not at all. Third, the validation panel wrote more comments and gave feedback on each item of the self-assessment instrument (survey). The panel addressed annotations, points of views and recommendations, including syntax, lexicon, and contents. The suggested editing and modifications were completed based on the panel's recommendations and suggestions. Fourth, a modified version of the self-assessment instrument was sent to the validation panel for additional review. The panel suggested more modifications regarding ILR terminologies and grammar. After that, the panel got the new modified version for additional evaluation. Finally, the panel approved the modified self-assessment instrument and it was ready to be tested for the reliability through "a test-retest" study which was done completed in a two-week period; a two-week period was chosen, because the examinees would not feel that they improved significantly over the course of two weeks of instruction. The panel who performed the evaluation, validation of the CDS, and reviewed the rules for Scoring Self-assessment survey where; Dr. Jackson Gordon: Research Specialist, Dr. Elfiky Salem: Oral Proficiency Interview (OPI) Specialist, and Dr. Boussalhi Abdelfattah: Testing Specialist.

Reliability of the Arabic reading ability self-instrument survey was established through a test-retest study with two alternative forms of the survey administered one to two week apart. In addition, reliability in terms of internal consistency of sub-scales in the assessment forms was analyzed using Cronbach's alpha. A group of ninety students studying Arabic at the DLIFLC participated in the test-retest study. The group consisted of thirty students in semester I, thirty students in semester II, and thirty students in semester III.

To minimize any rating bias and to protect human confidentiality, participants were asked not to write their names on the self-assessment forms. Each participant was given a code number from a master list. Each participant then wrote the code number, instead of their names on the forms. Testers scored both survey forms globally to provide two scores for each participant; used to check the reliability of the "parallel forms". A relationship analysis was used to examine the percentage of participants who received the exact same score on both forms A and B (exact agreement) in semester one, semester two, and semester three. In semester one, 29 (87.87 %) students scored the same level on Form A and Form B (exact agreement) of the survey. In semester two, 27 (84.37%) students scored the same level on Form A and Form B (exact agreement). In semester three, 28 (82.35%) students scored the same level on form A and form B (exact agreement). Finally, the survey reliability average among the three semesters Arabic students was measured at 84.86 %.

The researcher and the CDS validation panel then assigned numeric scores for each self-assessment question corresponding to each Arabic reading proficiency level. The numerical scores corresponded to each reading proficiency level codes in the survey

were (level 0+ = 06), (level 1 = 10), (level 1+ = 16), (level 2 = 20), (level 2+ = 26), and (level 3 = 30); these assigned numerical scores were based on how the Arabic self-assessment survey was designed, and the ILR language skill level descriptions of reading.

Rules for Scoring Self-Assessment

A specialist panel on testing protocols was formed to review the scoring rubrics of the CDS instrument. This panel was the same panel who validated the CDS instrument before. First, there was an individual meeting with the panel members to discuss the expectations and the roles of each member in the process of creating self-assessment scoring rules. Second, the self-assessment rules were given to the panel to write feedback and concerns for revision. Third, the testing specialist panel provided comprehensive and explicit feedback on the self-assessment rules. Next, more adjustments and corrections were made, based on the panel's feedback. Following that, the panel got a new update of the revised self-assessment rules based on their feedback, comments, and concerns for further modifications. Finally, the panel reexamined the revised self-assessment rules for additional changes and final approval. In order to score the students' responses, the five-point Likert scale was collapsed into three groups (A, B, and C), as in Table 1.

Table 1

Survey Items Scoring Rules

Group	CDS Response Choices	Students Reading Level
A	Quite Easily or Easily	Student is at the level of the can-do item
B	With some difficulty	Student is half a level below the level of the can-do item
C	With great difficulty or Not at all	Student is one level or more below the level of the can-do item

1. In order to be at a particular reading level, the student should respond to all CDS statements from that level by choosing “Quite Easily” or “Easily” (Group A).
2. The student is half a level below the level of the can-do item, if he/she selected even one of CDS statements “With some difficulty” (Group B); which means that the student is facing some difficulties and he/she is not quite at that level.
3. The student will go at least one whole level down, if he selected one of CDS statements with “With great difficulty” or “Not at all” (Group C); which means that the student is facing great difficulties and he/she is not at that level.
4. Students must meet the Reading low level requirements before the scorer moves to higher level items; i.e., scoring will stop as soon as the student’s responses indicate that he/she is not functioning at a given level. Scoring must start with Level 0+ questions.

5. The CDS contains 42 items and is based on the ILR Reading skill level descriptions of Arabic language. There are 2 items at level 0+, 7 items at level 1, 8 items at level 1+, 7 items at level 2, 10 items at level 2+, and 8 items at level 3.
6. Scoring Rules exception: the survey scorer will continue scoring if the student responded to one of the CDS statements (with one difficulty) for one time only, and the selected item/s that followed directly were(Quite Easily, or Easily).
7. Table 2 below explains how the reading level scores are derived from examinee responses to the can-do items.

Table 2

CDS Scoring Rules

Scoring Rules	Reading Levels	CDS Statements	CDS Student Responses		
R1	L0	N/A	All C of L0+	and B of L0+	
R2	L0+	(Statements 1-2)	All A of L0+ only	Or B of (L1)	Or C of (L1+)
R3	L1	(Statements 3-9)	All A of L1 only	Or B of (L1+)	Or C of (L2)
R4	L1+	(Statements 10-17)	All A of L1+ only	Or B of (L2)	Or C of (L2+)
R5	L2	(Statements 18-24)	All A of L2 only	Or B of (L2+)	Or C of (L3)
R6	L2+	(Statements 25-34)	All A of L2+ only	Or B of (L3)	
R7	L3	(Statements 35-42)	All A of L3 only		

Data Collection

Participation in the study was voluntary and it took approximately 40 minutes to complete the Arabic reading self-assessment survey. The study data collection was conducted during the school day in order to reduce the disturbance with students' military assignments. There were approximately 15 students who took the DLPT5 and graduated from the Arabic language program every two weeks; therefore, collecting data from students took about five months because it was based on their graduation timetable. The required data was the students' results from the CDS (self-assessment survey) of Arabic reading ability and the results from the Arabic DLPT5-Reading test administered one to two weeks later.

Prior to collecting data, a panel of language experts validated the CDS (self-assessment survey). After the CDS (self-assessment survey) was validated, a test-retest study was conducted with 90 students who took two alternative forms (A and B) of the CDS (self-assessment survey), administered one to two weeks apart to check the reliability of the instrument. The proof of the CDS reliability was the relationship between scores on the two instrument forms. The two forms of the self-assessment survey items are the same items but in a different order. The research study then moved to the final phase where 153 students took both the Arabic DLPT5-R and the CDS (self-assessment survey of Arabic reading ability).

Data Analysis

The following research questions and hypotheses guided the study:

RQ1: What is the correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

RQ2: What is the difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

RQ3: Is there a difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered?

H₁: There is a statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

H₀: There is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

The analysis of the above variables took place in two phases as recommended for correlational modeling studies (Babbie, 2012). First, the descriptive statistics were analyzed. Descriptive statistics of the dependent and independent variables were

summarized in terms of the frequency distribution and measures of central tendency (Bryman, 2012). In the frequency distributions, the number and the percentage of occurrence of the study variables were included. The measures of central tendency included the mean, standard deviation, median, minimum, and maximum values for the study variables. Descriptive statistics differ from inferential statistics in that descriptive statistics describe what the data set displays; whereas, inferential statistics draw conclusions about the population based-on the sample statistics (Plonsky & Gass, 2011). Graphical analysis using appropriate charts to present each variable were conducted as well. These graphs included frequency and pie charts.

The data for this study included the independent variable of the Arabic CDS scores (self-assessment survey) and the dependent variable of the Arabic DLPT5-Reading scores along with the control variables of the highest education level completed, military branch, military rank, gender, age, and previous experience with language learning. To address the first research question, the CDS scores of participants were analyzed to determine whether a relationship existed with the DLPT5-R scores. The CDS was determined to be valid if a statistically significant correlation exists with the DLPT5-R scores. A non-parametric test such as the Spearman's rho correlation analysis was used to calculate the correlation between the independent and the dependent measures because the CDS and DLPT5 scores represent ordinal data rather than interval data (e.g., 0, 0+, 1, 1+, 2, 2+). A correlation analysis is appropriate when the purpose of the analysis is to assess the relationship between two identified variables. If a significant correlation exists, then it can be concluded that there is sufficient evidence to reject the first null hypothesis. Moreover, the intraclass correlation coefficient (ICC) was used to determine

the strength of relationship within the same group. Since the CDS and the DLPT5 are non-interval in nature, it was appropriate to test whether scores within the same group resemble each other.

To address the second research question, it was necessary to compare the Arabic DLPT5-R scores of participants based on the CDS scores. Because both the DLPT5-R scores and the CDS scores are ordinal in nature, cross-tabs and Chi-square analysis were conducted to determine whether a difference exists between the Arabic DLPT5-R scores based on the CDS scores. A Chi-square analysis was utilized to determine significant differences between non-interval variables. This analysis considers the frequency of occurrence for each group of scores to determine whether difference exists. If a significant difference exists, then it can be concluded that there is sufficient evidence to reject the second null hypothesis.

For the third research question, an ordinal logistic regression analysis was conducted considering the variables of the highest education level completed, military branch, military rank, gender, age, and previous experience with language learning as control variables. An ordinal logistic regression rather than multiple linear regression was conducted because the dependent variable considered is ordinal in nature (Aiken, West & Pitts, 2008). Moreover, a bivariate analysis was conducted to examine potential confounding variables that should be considered in the regression analysis. The dependent variable was the Arabic DLPT5-R scores while the independent variable was the CDS scores. If a significant relationship exists ($p\text{-value} < .05$), then there is sufficient evidence to reject the third null hypothesis. A significance level of .05 was utilized for all statistical analyses.

Protection of Human Participants

The protection of human participants includes two key factors that the IRB weighs: informed consent and confidentiality. Eligible participants for the current study were provided an informed consent form (see Appendix F). Each form describes the rationale for the study, the premise of the study, and the intent of the study. The informed consent form also informed the participants that they may withdraw from the study at any time without reprisal or loss of benefit or penalty. Eligible participants were informed that the current study may be published in a nationally recognized peer-reviewed journal and that any personal information and the results of their particular surveys will be kept confidential. Potential participants were informed that participation in the current study poses no foreseeable risk for participating in the study. Prior to permitting eligible participants access to the paper-and-pencil survey, signed informed consent waivers were required.

Permission for the study was obtained from the Department of Defense (DoD) course administrators (see Appendix G). The permission included the ability to collect data from the DoD for the purpose of this non-invasive sociological study to validate a self-assessment instrument of Arabic reading ability and correlate it with the Arabic DLPT5-Reading score. To ensure confidentiality of the participants, both confidentiality and anonymity was enforced throughout the study. To further, protect the confidentiality of the study participants, the collected data was secured under password protection. The retention period of all surveys and documentation will be five years beginning on the date the current study was approved after submission. Deletion of the information will be completed only upon request by the participants should they wish their survey results be

removed from the study or after the study's completion. Only aggregate and statistical data from the study will be made available upon request. After the research is completed, the information (data) will be destroyed after a period of five years. At the end of the five years, all paper documents will be shredded, while electronic storage devices used for the survey and analysis process will be wiped clean and then physically destroyed.

Summary

This chapter provided the methodology that will be followed to conduct the quantitative correlational study. The choice of research design was discussed in this chapter along with a review of the purpose of the study. This study used the quantitative correlational method to determine the validity and reliability of the CDS (Arabic reading self-assessment instrument) with respect to the Arabic DLPT5-R scores.

The goal of the study was to provide statistical evidence of validity and reliability for the new CDS self-assessment instrument. The validated CDS instrument will be used as the baseline, and the control variables of the highest education level completed, military branch, military rank, gender, age, and previous experience with language learning are included. One DoD language-learning Arabic Basic Course was used to conduct the data collection targeting 153 participants. A minimum of 99 participants will be required to maintain a power level of 80%. Chapter 4 of the study provides the findings of the validity and reliability tests and Chapter 5 discusses the findings with respect to the relevant literature.

CHAPTER FOUR: FINDINGS

The objective of this quantitative study was to develop and validate a language self-assessment instrument of Arabic reading ability that can be used to obtain a reliable estimate of the Arabic reading proficiency test (DLPT5-R). The correlation between the two assessments: ratings obtained from the CDS (a self-assessment instrument) and scores obtained from the valid and reliable Arabic DLPT5-R Test of Reading ability was investigated to achieve the objective of the study. In line with this, this study was guided by the following research questions and hypotheses:

RQ1: What is the correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability.

RQ2: What is the difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability?

H₁: There is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

H₀: There is no statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability.

RQ3: Is there a difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered?

H₁: There is a statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

H₀: There is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

This chapter begins with the descriptive statistics analysis by summarizing the data through frequencies and percentages summary. This is followed by the results of the Spearman's rho correlation analysis, chi-square test, and ordinal logistics regression to address the research hypotheses presented.

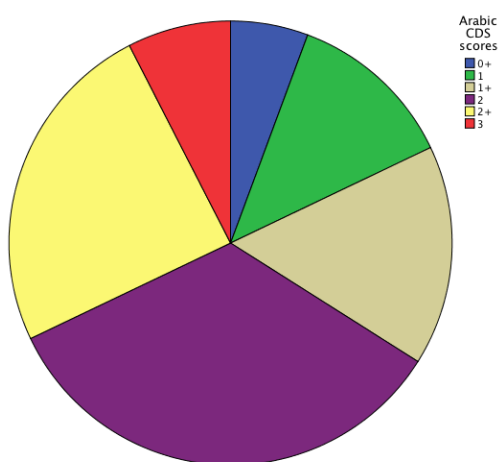
Results of Descriptive Statistics

A total of 107 Defense Department military students who were native speakers of English and who were learning Arabic as a second language participated in the study. For the scores of the self-assessment survey of CDS Arabic reading ability, more than half of the students have reading levels of 2 or 2+. Out of the 107 students, 36 (33.6%) students have reading levels of 2, and 26 (24.3%) students have reading levels of 2+. There were 8 (7.5%) who had reading levels of 3; while there were 6 (5.6%) who had reading levels of 0+. There were 13 (12.1%) students who had reading levels of 1, and there were 17 (15.9%) students who had reading levels of 1+ (see Table 3 and Figure 1).

Table 3

Arabic CDS Scores

	Frequency	Percent
0+	6	5.6
1	13	12.1
1+	17	15.9
2	36	33.6
2+	26	24.3
3	8	7.5
Missing	1	0.9
Total	107	100

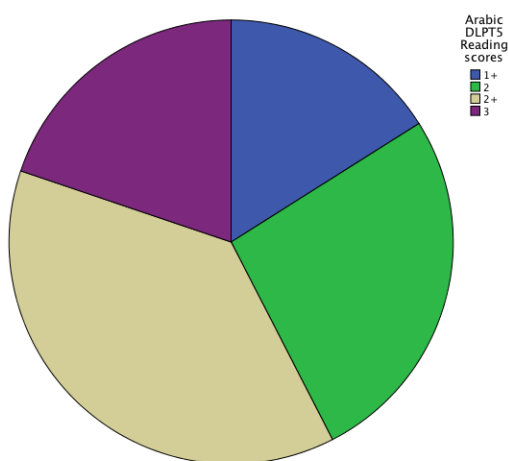
*Figure 1. Arabic CDS Scores*

For the reading levels of the students based on the DLPT5-R scores, more than half of the students have reading levels of 2 or 2+ or 3. Out of the 107 students, 40 (37.4%) students had reading levels of 2+, 28 (26.2%) students had reading levels of 2, and there were 21 (19.6%) who had reading levels of 3. There were also 17 (15.9%) students who had reading levels of 1+ (see Table 4 and Figure 2).

Table 4

Arabic DLPT5-R Scores

	Frequency	Percent
1+	17	15.9
2	28	26.2
2+	40	37.4
3	21	19.6
Missing	1	0.9
Total	107	100

*Figure 2. Arabic DPLT-5R Scores*

1. There were 23 (21.90 %) students scored the same level in the CDS and the DLPT5-R (e.g. an examinee CDS level was L1+ and his DLPT5 level was L1+).
2. There were 47 (44.33 %) students whose their scores were varied by a plus or minus half a level (+, -) between the CDS and the DLPT5-R (e.g. an examinee CDS level was L1 and his DLPT5 level was L1+ or vice versa).
3. There were 35 (33.01 %) students whose scores were varied by a plus or minus (+, -) one level or more between the CDS and the DLPT5-R (e.g. an examinee CDS level was L2 and his DLPT5 level was L0+ or L3).

For the highest education level completed by the Defense Department military students, 51 (47.7%) had completed high school, 34 (31.8%) had completed a Bachelor's degree, and 5 (4.7%) who had a Master's degree (see Table 5 and Figure 3). For the military branch, 66 students (61.7%) were Army, 17 (15.9%) were Navy, 15 (14%) were Air Force, and 8 (7.5%) were Marines (see Table 6 and Figure 4). The majority or 101 (94.4%) of the students had a military rank of enlisted and 5 (4.7%) were officer ranked (see Table 7 and Figure 5). In terms of gender, 72 (67.3%) were male and 34 (31.8%) were female (see Table 8 and Figure 6). Half of the students or 60 (56.1%) had an age in the range between 21 to 25 year old, 21 (19.6%) had an age in the range of 26 to 30 years old, 12 (11.2%) had an age in the range of 31 to 35 years old, and 11 (10.3%) had an age in the range of 18 to 20 years old (see Table 9 and Figure 7).

Table 5

Highest Education Level Completed

	Frequency	Percent
GED	3	2.8
HS	51	47.7
AA	13	12.1
BA	34	31.8
MA	5	4.7
Missing	1	0.9
Total	107	100

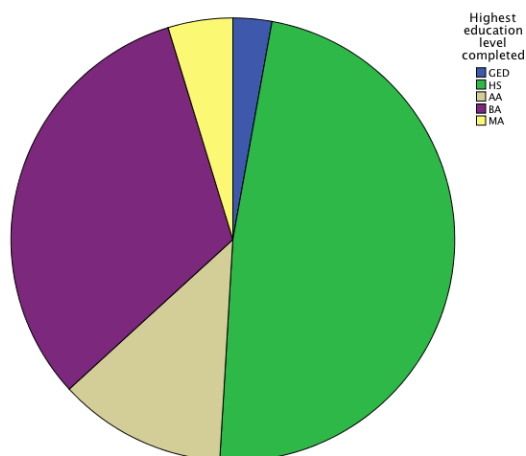


Figure 3. Highest Education Level Completed

Table 6

Military Branch

	Frequency	Percent
Army	66	61.7
Air Force	15	14
Navy	17	15.9
Marines	8	7.5
Missing	1	0.9
Total	107	100

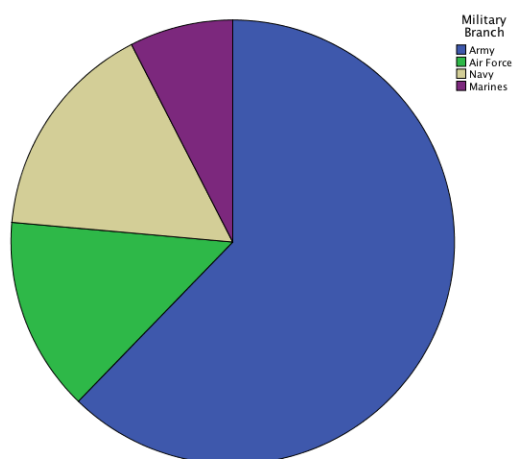


Figure 4. Military Branch

Table 7

Military Rank

	Frequency	Percent
Officer	5	4.7
Enlisted	101	94.4
Missing	1	0.9
Total	107	100

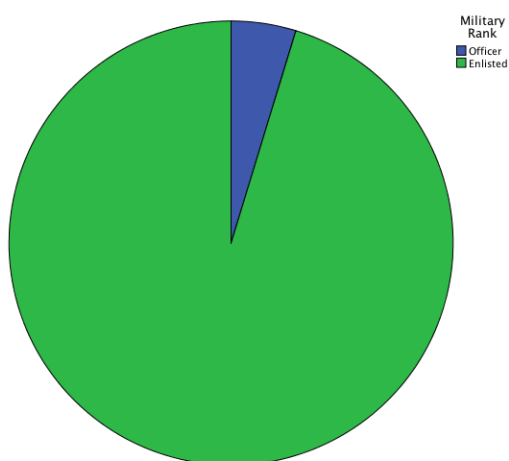
*Figure 5. Military Rank*

Table 8

Gender

	Frequency	Percent
Male	72	67.3
Female	34	31.8
Missing	1	0.9
Total	107	100

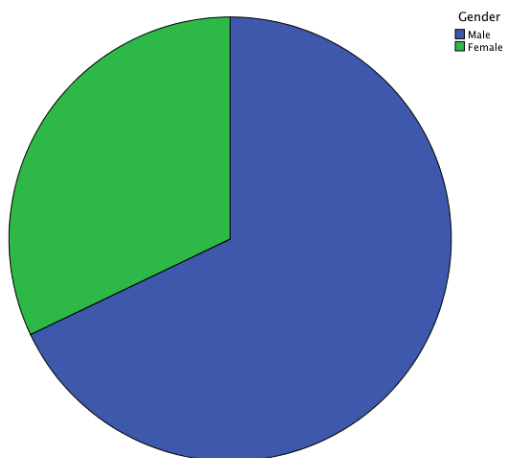


Figure 6. Gender

Table 9

Age

Age	Frequency	Percent
18-20	11	10.3
21-25	60	56.1
26-30	21	19.6
31-35	12	11.2
36-40	1	0.9
Missing	2	1.9
Total	107	100

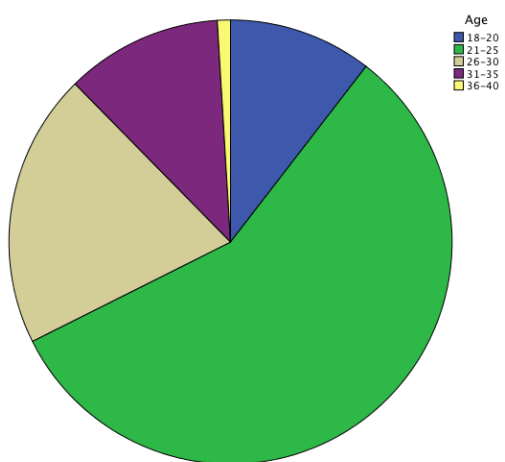


Figure 7. Age

The majority of the students (90; 84.1%) had no previous experience with reading a language other than English (see Table 10 and Figure 8). Also, the majority of the students had no previous language experience other than English (90; 84.1%). Of those with prior language experience 8 or 7.5 % of the students had this experience in reading Spanish (see Table 11). In addition to that, 3 or 2.8 % of the students first started reading English when they were 1 years old, 2 or 1.9% of the students first started reading English when they were 8 years old, and 2 or 1.9% of the students first started reading English when they were 10 years old (see Table 12 and Figure 9).

Table 10

Previous Experience Reading a Language Other Than English

	Frequency	Percent
Yes	16	15
No	90	84.1
Missing	1	0.9
Total	107	100

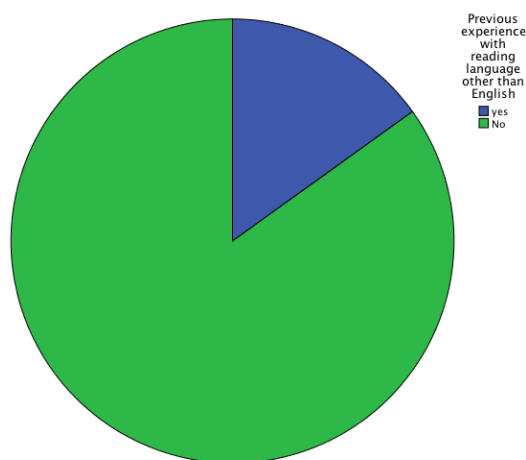


Figure 8. Previous Experience Reading a Language Other Than English

Table 11

Previous Language Reading Experience Other Than English

Language	Frequency	Percent
Cambodian	1	0.9
German	2	1.9
Korean	1	0.9
Nyanja	1	0.9
Spanish	8	7.5
Total	107	100

Table 12

Age First Started Reading English

Age	Frequency	Percent
1	3	2.8
2	1	0.9
3	1	0.9
4	1	0.9
5	1	0.9
6	1	0.9
8	2	1.9
10	2	1.9
Missing	95	88.8
Total	107	100

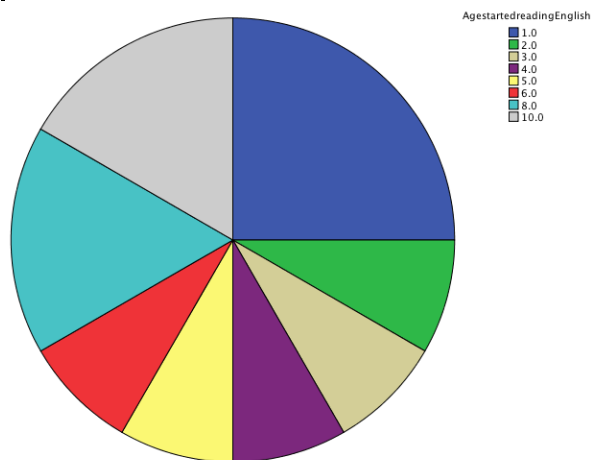


Figure 9. Age First Started Reading English

Also, 11 out of the 107 students (10.3%) had previous experience studying Arabic before coming to DLIFLC (see Table 13 and Figure 10). In addition to that 2 or 1.9 % of the students studied Arabic for 2 years before studying at the DLIFLC. Also 2 or 1.9 % of the students studied Arabic for 1 year before studying at the DLIFLC (see Table 14 and Figure 11). Most of the students (79; 73.8%) had previous experience studying any other foreign language (see Table 15 and Figure 12). Almost half of these students with prior study experience, had studied Spanish (45.2%), 11.3% studied French, and 12.2% studied German (see Table 16). The duration of prior studying of another foreign language, ranged between 1 to 24 years but most of the students studied the foreign language between the range of 1 and 4 years (see Table 17).

Table 13

Previous Experience Studying Arabic Before DLIFLC

	Frequency	Percent
No	95	88.8
Yes	11	10.3
Missing	1	0.9
Total	107	100

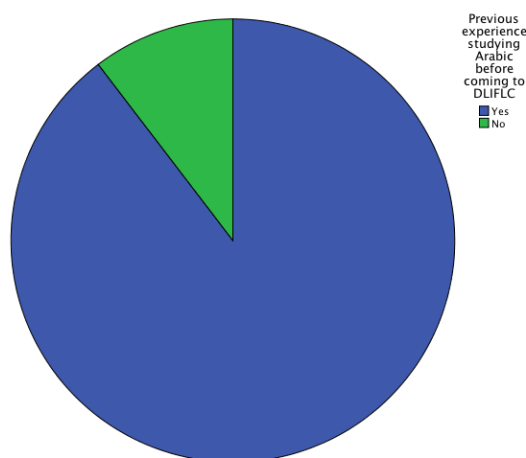


Figure 10. Previous Experience Studying Arabic Before DLIFLC

Table 14

Length of Time Studying Arabic Before DLIFLC

Time	Frequency	Percent
1 month	1	0.9
1 year	2	1.9
1 year 2 months	1	0.9
1 year 6 months	1	0.9
2 years	2	1.9
3 months	1	0.9
5 months	1	0.9
6 months	1	0.9
9 months	1	0.9
Total	107	100

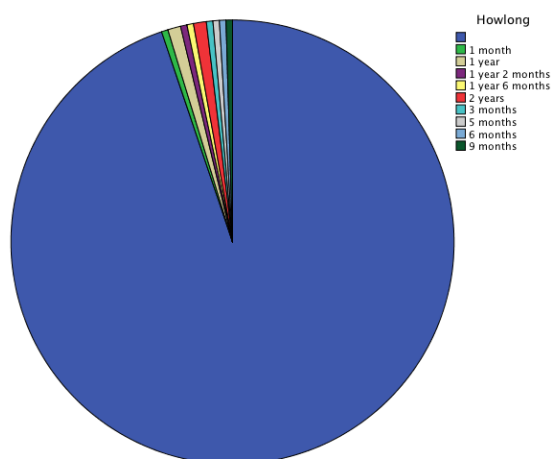


Figure 11. Length of Time Studying Arabic Before DLIFLC

Table 15

Previous Experience Studying Other Foreign Languages Before DLIFLC

	Frequency	Percent
Yes	79	73.8
No	27	25.2
System	1	0.9
Total	107	100

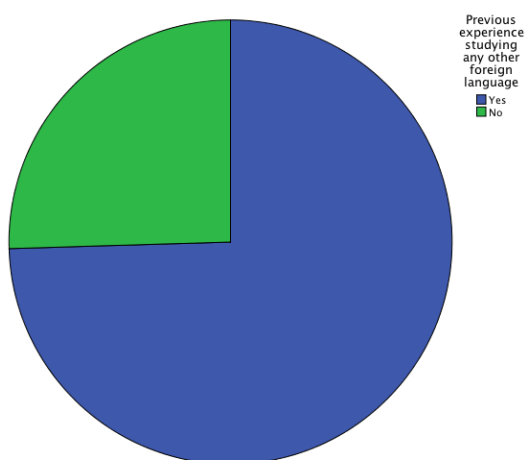


Figure 12. Previous Experience Studying Other Foreign Languages Before DLIFLC

Table 16

Previously Studied Other Foreign Language Before DLIFLC

Language	Frequency	Percent
Ancient Greek	1	0.9
Cambodian	1	0.9
Cambodian/Khmer	1	0.9
Chinese	5	4.3
Czech	1	0.9
French	13	11.3
German	14	12.2
Greek	1	0.9
Hebrew	2	1.7
Italian	4	3.5
Japanese	2	1.7
Korean	3	2.6
Latin	4	3.5
Latvian	1	0.9
Nyanja	1	0.9
Pashto	1	0.9
Portuguese	1	0.9
Russian	3	2.6
Slovak	1	0.9
Spanish	52	45.2
Spanish 3 years	1	0.9
Tagalog (Filipino)	1	0.9
Uzbek	1	0.9
Total	115	100

Table 17

Length of Time Studying Other Foreign Language before DLIFLC

Time	Frequency	Percent
1 year	21	18.3
1 year 2 months	1	0.9
1 year 4 months	1	0.9
1 year 6 months	2	1.7
2 years	27	23.5
2 years 6 months	1	0.9
23 years	1	0.9
25 years	1	0.9
3 years	16	13.9
3 years 4 months	1	0.9
3 years 6 months	1	0.9
4 year	1	0.9
4 years	13	11.3
5 months	1	0.9
5 years	6	5.2
6 months	9	7.8
6 years	5	4.3
6 years 6 months	1	0.9
7 years	1	0.9
8 months	1	0.9
8 years	2	1.7
9 years 7 month	1	0.9
Total	115	100

Correlation Results for the Arabic CDS and the Arabic DLPT5-R Test

Table 18 summarizes the results of the Spearman's rho correlation analysis to determine the relationship between CDS and DLPT5-R scores. The results showed that the Arabic CDS scores and Arabic DLPT5 Reading scores ($r = 0.13$, $p = 0.19$) were not significantly correlated since the p-values were greater than the level of significance value of 0.05. Thus, the results failed to reject the null hypothesis for research question one. There is no correlation between the Arabic DLPT5 test in reading and the self-

assessment survey of Arabic reading ability.

Table 18

Spearman's Correlation Result Between Arabic CDS Scores and Arabic DLPT5-R

			Arabic DLPT5 Reading scores
Spearman's rho	Arabic CDS scores	Correlation Coefficient	0.13
		Sig. (2-tailed)	0.19
		N	106

A scatter plot in Figure 13 was generated to graphically show the correlation between the Arabic CDS scores and Arabic DLPT5-R scores. The scatter plot shows that there was no line pattern which supports the result of the Spearman's correlation test that there is no correlation between CDS scores and Arabic DLPT5 Reading scores. A significant correlation would show a linear pattern in the scatter plot between the two variables.

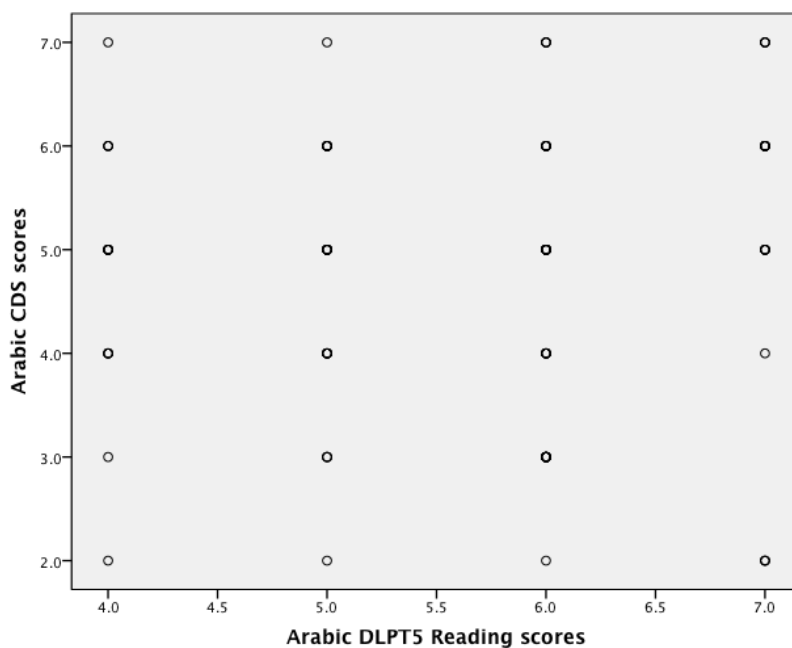


Figure 11. Scatter Plot of Arabic CDS Scores and Arabic DLPT5 Reading Scores

The intraclass correlation coefficient (ICC) between the Arabic CDS scores and Arabic DLPT5 Reading scores was investigated to determine the strength of relationship of the two similar scores within the same group. The ICC statistics is used to determine the strength of the correlation of data in the same group that resemble each other. The statistic measures the reliability or consistency of the scores of different measures. Typically, this is used when there are many respondents answering a survey to determine whether the responses of the different responses were consistent with each other. The statistical result ($f[105, 105] = 1.18, p = 0.20$) was insignificant since the p-values were greater than the level of significance value of 0.05. The results showed that the Arabic CDS scores and Arabic DLPT5 Reading scores within the same group, and did not resemble each other. Participants have different scores in Arabic CDS scores and Arabic DLPT5 Reading scores. The two scores did not resemble each other since the intraclass correlation coefficient for both the single measure ($\alpha = 0.08$) and the average measures ($\alpha = 0.15$) also showed that the values were extremely low. The ICC statistics could range between 0 and 1 and higher values would indicate a higher correlation between the two scores.

Table 19

Intraclass Correlation Coefficient Results

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0.08 ^a	-0.11	0.27	1.18	105	105	0.2
Average Measures	0.15 ^c	-0.24	0.42	1.18	105	105	0.2

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Chi-Square Test Results

Chi-Square test was conducted to determine whether a difference existed between the Arabic DLPT5-R scores based on the CDS scores. The chi-square test result is summarized in Table 20. The results showed that there was a significant difference in the Arabic reading scores in the CDS test and DLPT5 Reading test ($\chi^2 (5) = 28.80, p = 0.00$). There was a significant difference since the p-value was less than the level of significance value of 0.05. The null hypothesis for research question 2 was rejected. The results supported the alternative hypothesis that there is a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability. The cross tabulation in Table 21 shows that there was significant difference in the reading levels of 0, 0+, and 3 in the CDS test and DLPT5 Reading test. There were more students that have 0 and 0+ reading levels in the CDS test than in the DLPT5-r test, which have no students scoring 0 or 0+. On the other hand, there were significantly lesser students that have reading level of 3 in the CDS test (8)

than in the DLPT5-R test (21). It was also observed that there were lesser students that have reading level of 2+ in the CDS test (26) than in the DLPT5-R test (40) while there were significantly greater students that have reading level of 2 in the CDS test (36) than in the DLPT5-R test (28).

Table 20

Chi-Square Test Results

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	28.80 ^a	5	0
Likelihood Ratio	36.375	5	0
Linear-by-Linear Association	22.507	1	0
N of Valid Cases	212		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 3.00.

Table 21

Cross Tabulation Results Between Arabic Reading Scores

Reading Scores	CDS	DLPT5-R	Total
0	6	0	6
0+	13	0	13
1+	17	17	34
2	36	28	64
2+	26	40	66
3	8	21	29
	106	106	212

Results of Ordinal Logistic Regression

The result of the ordinal logistic regression analysis to address research question three was summarized in Table 22. The analysis determined if the Arabic DLPT5-R scores and the CDS are significantly different when the variables of the highest education level completed, military branch, military rank, gender, age, and previous experience with language learning were introduced as control variables to determine the potential confounding effect of these variables in the relationship between the independent variable and dependent variable. The dependent variable in the regression was the Arabic DLPT5-R scores while the independent variable was CDS scores. The results showed that all the control variables of highest education level completed (Wald [1] = 0.07, $p = 0.79$), military branch (Wald [1] = 1.31, $p = 0.25$), military rank (Wald [1] = 0.01, $p = 0.92$), gender (Wald [1] = 1.90, $p = 0.17$), age (Wald [1] = 0.00, $p = 0.98$), previous experience with reading language other than English (Wald [1] = 2.99, $p = 0.08$), previous experience studying Arabic before coming to DLIFLC (Wald [1] = 0.01, $p = 0.92$), and previous experience studying any other foreign language (Wald [1] = 0.25, $p =$

0.62) did not have a significant confounding effect since the p-values were all greater than the level of significance value of 0.05. In addition, each of the ordinal values of the independent variable of CDS scores of reading level 0+ (Wald [1] = 0.12, $p = 0.73$), 1 (Wald [1] = 0.90, $p = 0.34$), 1+ (Wald [1] = 3.62, $p = 0.06$), 2 (Wald [1] = 3.18, $p = 0.07$), and 2+ (Wald [1] = 0.21, $p = 0.625$) were insignificant. The results failed to reject the null hypothesis for research question three. The results showed that there is no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

Table 22

Results of Ordinal Logistic Regression Test

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[ArabicDLPT5Readingscores = 1+]	-2.65	2.96	0.8	1	0.4	-8.46	3.16
	[ArabicDLPT5Readingscores = 2]	-1.16	2.96	0.16	1	0.7	-6.96	4.64
	[ArabicDLPT5Readingscores = 2+]	0.73	2.95	0.06	1	0.8	-5.06	6.52
Location	Highest Education Level Completed	0.05	0.2	0.07	1	0.8	-0.34	0.45
	Military Branch	-0.23	0.2	1.31	1	0.3	-0.63	0.17
	Military Rank	-0.11	1.08	0.01	1	0.9	-2.22	2.01
	Gender	-0.57	0.41	1.9	1	0.2	-1.38	0.24
	Age	-0.01	0.27	0	1	1	-0.53	0.51
	Previous experience with reading language other than English	0.93	0.54	2.99	1	0.1	-0.13	1.99
	Previous experience studying Arabic before coming to DLIFLC	-0.07	0.63	0.01	1	0.9	-1.3	1.17
	Previous experience studying any other foreign language	-0.22	0.43	0.25	1	0.6	-1.06	0.63
	[ArabicCDSscores=0+]	-0.38	1.1	0.12	1	0.7	-2.53	1.77
	[ArabicCDSscores=1]	-0.83	0.88	0.9	1	0.3	-2.55	0.89
	[ArabicCDSscores=1+]	-1.61	0.85	3.62	1	0.1	-3.27	0.05
	[ArabicCDSscores=2]	-1.37	0.77	3.18	1	0.1	-2.87	0.14
	[ArabicCDSscores=2+]	-0.36	0.78	0.21	1	0.7	-1.88	1.17
	[ArabicCDSscores=3]	0 ^a	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Summary

This chapter provided the results of the statistical analysis to address the research questions of the quantitative correlational study. For research question one, the results of the Spearman's rho correlation analysis showed that there was no correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability. For research question two, the results of the Chi-square test showed that there was a statistically significant difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability. For research question three, the results of the ordinal logistic regression showed that there was no statistically significant difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered.

CHAPTER FIVE: DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Introduction

Chapter 5 summarizes the entire dissertation and discusses its findings and its implications on the development of a language self-assessment instrument in Arabic proficiency. The study expanded current literature since the relationship between student's Arabic reading self-assessment and ratings obtained from a formal Arabic DLPT5 in the reading skill had yet to be explored. The chapter begins by presenting an overview of the study and then discusses the purpose and significance of the topic followed by the enumeration of the three research questions. Then, the results of the CDS and its correlation to the scores on the DLPT5 reading tests are discussed in relation to existing research and the implications of the findings on foreign language learning in the U.S. Next, the researcher offers recommendations to expand the current study or generalize the results of future studies study before finally making a conclusion.

Overview of the Study

The NSLI highlighted the need for the U.S. to increase its capacity to provide experts with critical language skills that are vital to national security and foreign policy (O'Connell & Norwood, 2007). Foreign language proficiency and knowledge of cultures increases global competencies to help Americans meet the demands of a global workforce (Goodman, 2012; Ochoa, 2012). Since every sector of the U.S. depends on language services, it is imperative to increase knowledge on foreign languages and cultures to preserve U.S. security and the economy (O'Connell & Norwood, 2007; Fenstermacher, 2012). At the secondary level, only 30% of students in the U.S. are studying foreign languages, significantly lower than the 59.6% of students studying two

or more languages within the E.U. at a comparable educational level (EuroStat, 2012). For the DoD, Junor (2012) noted that personnel with the required language proficiency level are only at 28%, or 10,377, out of a total of 36,983 military language positions. The remaining positions are filled with personnel that do not have the required language proficiency level. If the DoD is able to meet the qualified personnel requirements, the department will be in a better position to strengthen its relationship with its allies, remain engaged in the international arena, and continue communicating with local people and senior officials to ensure U.S. security.

The study seeks to address the language proficiency gap in the DoD by identifying the relationship between a student's Arabic reading self-assessment and ratings obtained from a formal Arabic DLPT5 reading test. Currently, there are no studies that examine that relationship. Additionally, the effects of demographic variables on self-assessment and language proficiency were also briefly discussed. The relationship was determined by the answers to three research questions:

RQ1: What is the correlation between the Arabic DLPT5 test in reading and the self-assessment survey of Arabic reading ability?

RQ2: What is the difference in scores between the Arabic DLPT5 test scores in reading and the self-assessment survey of Arabic reading ability?

RQ3: Is there a difference in scores on the Arabic reading self-assessment survey and the Arabic DLPT5-R test when the control variables are considered?

The null hypothesis for all three research questions imply no statistically significant difference while the alternate hypothesis states that a statistically significant difference exists. The development of a language self-assessment instrument that can

obtain a reliable estimate of Arabic reading proficiency may be used by teachers, students, and U.S. soldiers at the DLIFLC to monitor and improve their required language proficiency level.

In order to answer the main research questions, the study utilized a quantitative correlational research study that investigated the correlation between two assessments, namely, ratings obtained from the CDS and scores obtained from a valid and reliable Arabic DLPT5 reading test. Spearman's rho, a non-parametric test, was used in the analysis to determine the degree of correlation between the independent and dependent variables. A total of 107 U.S. male and female military students from the four branches of the service participated in the study. They were all native speakers of English and were learning Arabic as their second language for 63 weeks in Monterey, California, aged 18 to 40, learn in the same classrooms, and have the same Arabic language program. The only variations are in terms of work experience, educational degrees, educational experiences, experiences in language learning, and educational attainment. The theories that explained the underlying fundamentals behind self-assessment and testing include the theories of Constructive Learning, Multiple Intelligences, and Social Cognitive.

Summary of the Results

Two tests and one regression analysis were conducted to decide whether the null hypothesis for all three main research questions should be rejected. For the first question, a Spearman's rho correlation analysis determined the relationship between CDS and DLPT5 reading scores. The results showed that both were not significantly correlated ($r = 0.13$, $p = 0.19$) since the p-values was greater than the significance level of 0.05. Therefore, the null hypothesis was not rejected and it was established that there is no

correlation between the Arabic DLPT5 reading test scores and the self-assessment survey of Arabic reading ability. A scatter plot was conducted and showed that no line pattern existed, verifying the results of the Spearman's rho correlation test. The ICC statistics between the Arabic CDS scores and the Arabic DLPT5 reading scores found that neither score within the same group resembled each other ($f(105, 105) = 1.18, p = 0.20$).

The second research question was answered through a chi-square test to determine whether a difference existed between the Arabic DLPT5 reading and CDS scores. The findings indicated that there was a significant relationship in the Arabic reading scores in the CDS test and the DLPT5 reading test ($\chi^2(5) = 28.80, p = 0.00$) since the p-value was less than the significance level of 0.05. Therefore, the alternative hypothesis was supported which showed a statistically significant difference in the scores between the Arabic DLPT5 reading and the CDS scores. A cross tabulation illustrated that there were more students who scored 0 and 0+ in the CDS test as compared to the DLPT5 reading test. On the other hand, a significantly higher number of students scored a 3 in the DLPT5 reading score as compared to the CDS score.

An ordinal logistic regression tried to answer the third research question. The results showed that all the control variables did not have a significant compounding effect and all the ordinal values of the CDS scores were insignificant. Therefore, the null hypothesis of no statistically significant difference in the scores when the control variables were considered was not rejected.

Discussion of the Results in Relation to Literature

The study's main contribution to existing literature is analyzing the effect of demographics on and the relationship between self-assessment scores and a formal

Arabic reading test scores. Most of current studies examined the effects of self-assessment on learning the English language while none have tackled Arabic. The literature review provided several explanations based on the three theoretical foundations on the importance of self-assessments for students. For the social cognitive theory, Woolfolk (2007) mentioned the need for self-regulation in helping students get feedback from teachers and in constructing and building their knowledge and skills. Meanwhile, Grabe (2009) argued that self-perception is hinged on the feeling of self-efficacy which is a good indicator of learning, motivation, and achievement. Meanwhile, the constructive learning theory explained how learners use resources, information, or assistance from others, experiences, and problem-solving strategies, and their mental ability in constructing new knowledge (Woolfolk, 2007). Lastly, the multiple intelligences theory stated that humans have unique intelligences that vary from person to person. Therefore, it is important for students to be given the chance to be active self-assessors, self-monitor their achievements, manage what they learn, and in critiquing their weakness and strengths, and to recognize the process of how they learned, and what they need to learn to achieve their goals. Through these methods, students are able to create their own understanding and perception of subjects that they learn (Campbell, 1999).

Based solely on the results of the study, the CDS can still be improved to become an adequate assessment tool of the Arabic language proficiency of military students. For the first research question, the presence of no significant correlation between the CDS and DLPT5 reading scores means that the both scores are not related. Therefore, a student's score on a CDS does not necessarily imply the same score for the DLPT5. The finding is contrary to the study of Elfiky (2012) which found that a significant correlation

existed between CDS and OPI. The percentage of agreement between both tests was also shown to be 58%. The results also contradict the conclusions of Woo (1995) where a student's self-assessment of language proficiency is the best predictor of DLPT III in learning Korean and Wolochuk (2009) where a significant correlation existed between self-assessment and the TOEFL. However, the findings agree the conclusions of Brantmeier et al. (2012) that the self-assessment of reading ability as a second language was not a correct predictor of a reading test performance. Yuko and Lee (2010) also agree because since their study showed a minimal positive effect of self-assessment on students learning English as a foreign language in South Korea. Despite this shortcoming, Brantmeier et al. (2012) noted that a benefit of self-assessment is being able to document the performance and learning of students over time.

Chen (2008) found out that feedback, training, and practicing self-assessment increased the accuracy of student's self-assessments because it helped students achieve their learning objectives and goals. However, the significant difference between the CDS and DLPT5 scores hinted a divergence of assessments as the CDS scores are significantly different from DLPT5 scores at a majority of the language proficiency levels. Students are actually underestimating their capabilities because they most rate themselves lower than their actual scores in the exam. Blanche (1988) offer an explanation that weaker students usually overestimate themselves as compared to higher achiever students. Since 87 percent of the participants have experience learning another language while 96 percent have graduated at least from high school, it can be deduced that the participants have the proper training and experience to overachieve and thus have the tendency to underestimate their own capabilities. The idea becomes more apparent as only eight

participants assessed themselves as Level 3 in the CDS but in reality, 21 students were rated Level 3. Other students rated themselves as 0 or 0+ while none achieved those scores during the DLPT5 reading test.

The answer to the second research question showed a significant difference in scores highlighting the probability that self-assessment tests have different impacts on different stages of learning a foreign language. Similar to the argument of Blanche (1988), self-evaluation or self-assessment is important during the early stages of studying and promoting the acquisition of a new language. Since self-assessment also motivates students (Blanche, 1988; Jiang 1999), the assimilation of a new language might become easier for students especially for novices or those that scored 0 or 0+. Additionally, Morgan (1985, as cited in Sternberg, 2002) found that students who self-monitor their goals and kept track of their progress were more successful, studied more, and scored better than students who did otherwise.

A significant difference between the CDS score and the DLPT5 Arabic reading score was determined by the study. However, the difference cannot be attributed to demographic variables since the control variables were deemed insignificant as noted for the third research question. The significant difference is similar to the findings of other researchers who examined the effects of self-assessments in learning English under various settings (Baniabdelrahman, 2010; LeBlanc & Painchaud, 1985; Wan-a-rom, 2008). Yuko and Lee (2010) offered an explanation on the difference of scores since the control variables were all insignificant. The authors posit that teachers and students look at self-assessment differently, thus the effectiveness depends on how both stakeholders look at the context of teaching and learning and how teachers view assessment.

Synthesizing the results of the two tests and one ordinal logistic regression, it can be inferred that a difference does exist between a military student's CDS score and DLPT5 Arabic reading score, but both are not correlated. In addition, these differences are not attributable to certain control variables examined in the study. The CDS may not be the most appropriate tool at the moment to monitor and improve the required language proficiency level of teachers, students, and military students since it does not serve as an indicator of overall exam performance. However, it should not be discounted as a tool to help improve the scores of students. Campbell (1999) and Grabe (2009) proposed several methods of reading assessments that may be considered by military foreign language educators. Since the literature is abundant with studies highlighting the benefits of self-assessments, other self-assessment methods may be used to supplement the results of the CDS in order to better document the progress of military student language learning.

Limitations

Chapter One presented four limitations that bounded the scope and results of the study. The first limitation restricted the applicability of the results to other institutions, universities, or bilingual programs because the subjects were military linguists. Besides the target population, the second limitation constrained the applicability of results to other language programs that are not using ILR skill level descriptions in evaluating students. Other individuals with various backgrounds and other different programs may provide a different set of results. It is important to understand the effects of the evaluation method and the background of the individual on the assessment results. Additional studies are recommended to shed light on this relationship and shall be subsequently discussed. Meanwhile, participant dropout also posed a risk, as explained

by the third limitation. It is possible for participants to withdraw from the study because they could not take the DLPT5-Reading test or may be asked to accomplish other duties. The researcher mitigated this concern by increasing the number of participants to 107, 8 more than the minimum required to make an assessment power level of 80 percent. Finally, the last limitation pertained to the survey instrument. The CDS instrument was written entirely in English and may not be understood by individuals who do not speak English as their first language. To mitigate this, the researcher only chose native speakers of English as the participants.

Implication of the Results for Practice

Military educators and even foreign language teachers should consider the results of this study in assessing the level of foreign language learning of students. The findings hinted that the CDS may not be the best or suitable assessment tool in determining the language proficiency of students. Although a self-assessment, based on theory, is important, other factors that were not considered in study might play a role in the variance between the self-assessment score and the actual score on the exam. Therefore, the findings underscore the challenge for these professionals to devise ways or plans to properly assess students in their language proficiency. Since a significant difference was found out for those that scored 0, 0+, and 3 in their CDS, the self-assessment might be too conservative for students since they would rank themselves lower than their actual level of proficiency. Although it cannot be discounted that personalities might influence the self-assessment, the CDS must be continuously improved in order to ensure that it becomes an adequate indicator of the student's scores in the actual exams. Some

suggested improvements could be using a similar sentence construction, syntax, points of view, and the like as compared to the DLPT5 reading score.

Campbell (1999) also offered several methods for self-assessment besides the usual standardized exam that may be utilized by military educators which include writing journals, portfolios, peer assessments, informal student/teacher dialogue, and self-reflection sheets. Also, educators could consider immersing the students in environments that will force them to use the foreign language. Trips to Middle Eastern countries where students can test their Arabic reading and speaking proficiency before taking the assessment or DLPT5 exam could improve the scores for both tests. Besides just learning, and possibly just memorizing, the language, it is highly important for the student to be able to communicate using the foreign language.

Recommendations for Further Research

The scope and limitations of the study have been restricted given the focus of the study on foreign language proficiency levels in the DoD. It would be insightful for future researchers to widen the scope of the study, examine other departments or institutions, or change the composition of the participants to contribute to the wealth of knowledge on foreign language proficiency gap in the U.S. At this point, the researcher would like to recommend the following expansions or topics:

1. Consider increasing the sample size. The study failed to reject the majority of the null hypotheses, so an increase in the number of participants might improve the outcome of future studies. A broader data base would be able to provide a more accurate picture of the relationship between a self-assessment score and the DLPT Arabic reading score.

2. Drill down the effects depending on Arabic reading proficiency level. It is possible that the impact of the self-assessment could be more significant when the student is still a novice as compared to an intermediate, advanced, or superior level of adeptness. The results could be used to match the appropriate assessment tool based on the level of proficiency of the student.
3. Analyze the impact of the tests on students that are not native English speakers or from educational institutions. Although the results showed that demographics did not play a role in determining a difference in the scores of the students, it might be interesting to see whether a student's language background and environment can influence a difference in scores.
4. Investigate the effects of other self-assessment methods on foreign language proficiency. A standardized test might not be the best form of assessment tool for language proficiency. Further studies may look at the various methods presented in literature and determine its effect on language proficiency. Several methods may also be combined to see its overall effect of language test scores.

Conclusion

The U.S. needs to increase its capacity to provide experts with critical language skills that are vital to national security and foreign policy. Junor (2012) noted that DoD personnel with the required language proficiency level are only at 28% with the remaining positions filled with personnel that do not have the required language proficiency level. If the DoD is able to meet the qualified personnel requirements, the department will be in a better position to strengthen its relationship with its allies, remain engaged in the international arena, and continue communicating with local people and

senior officials to ensure U.S. security. The study seeks to address the language proficiency gap in the DoD by identifying the relationship between a student's Arabic reading self-assessment and ratings obtained from a formal Arabic DLPT5 reading test and the effect of demographic variables on the relationship, if it exists. The study utilized a quantitative correlational research study on 107 U.S. male and female military students.

The Spearman's rho correlational test showed no correlation between the CDS score and the DLPT5 Arabic scores of students. Meanwhile, the chi square test determined that a significant difference exists between the CDS and DLPT5 scores but the ordinal logistic regression found out that the difference could not be attributed to the demographic characteristics of the participants. Therefore, the results imply that the CDS might not be an appropriate tool in assessing military student language proficiency and recommendations for practitioners were subsequently discussed. Further research is suggested to consider increasing the sample size, drill down the effects of self-assessment per proficiency level, analyze the impact of the test on other student populations, and investigate the effects of other assessment methods.

REFERENCES

- Aiken, L. S., West, S., & Pitts, S. C. (2008). Multiple linear regression. *Handbook of Psychology*. doi: 10.1002/0471264385.wei0219.
- Akaka, K. D. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- Armstrong, T. (2009). *Multiple intelligences in the classroom*. (3rd ed.). Alexandria, VA: Association for Supervision & Curriculum.
- Brindley, G.(2001). *The Cambridge Guide to Teaching English to Speakers of Other Languages*. Retrieved from: <http://libproxy.edmc.edu/login?qurl=http%3A%2F%2Fwww.credoreference.com/entry/cupteacheng/assessment>
- Babbie, E. R. (2012). *The practice of social research*. Belmont, CA: Wadsworth.
- Balnaves, M. & Caputi, P. (2001). *Introduction to quantitative research methods: An investigative approach*. Thousand Oaks, California: SAGE.
- Baniabdelrahman, A. A. (2010). The effect of the use of self-assessment on EFL students' performance in reading comprehension in English. *The Electronic Journal for English as a Second Language*, 14(2).
- Bannock, G., Davies, E., Trot, P., & Uncles, M. (2003). *New product development*. In The New Penguin Business Dictionary. Retrieved from http://libproxy.edmc.edu/login?qurl=http%3A%2F%2Fwww.credoreference.com/entry/penguinbus/new_product_development.
- Beerkens, R. (2010). *Receptive multilingualism as a language mode in the Dutch-German border area*. Münster, Germany: Waxmann.
- Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *REL C: A Journal of Language Teaching and Research*, 19(1), 75-93.
- Brantmeier, C., Vanderplank, R., & Strube, M. (2012). What about me? Individual self-assessment by skill and level of language instruction. *System*, 40(1), 144-160. doi:10.1016/j.system.2012.01.003
- Bryman, A. (2012). *Social research methods* (4th ed.). Oxford, England: Oxford University Press.

- Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge, England: Cambridge University Press.
- Butler, Y., & Jiyeon, L. (2006). On-Task Versus Off-Task Self-Assessments Among Korean Elementary School Students Studying English. *Modern Language Journal*, 90(4), 506-518. doi:10.1111/j.1540-4781.2006.00463.x
- Byers, L. M. (2010). *I know "I can": A validity study of a foreign language self-assessment*. The University of Tennessee at Chattanooga). *ProQuest Dissertations and Theses*, 60. Retrieved from <http://search.proquest.com/docview/851890288?accountid=34899>. (851890288).
- Campbell, L., Campbell, B., & Dickinson, D. (1999). *Teaching & learning through multiple intelligences*. (2nd ed.). Boston, MA: Allyn & Bacon.
- Chen, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262.
- Defense Language Institute Foreign Language Center [DLIFLC]. (2012). *Defense Language Proficiency Testing System 5 Framework 2012*. Retrieved from www.dliflc.edu/file.ashx?path=archive/documents/Framework
- Defense Language Institute Foreign Language Center [DLIFLC]. (2013). *Mission statement*. Retrieved from <http://www.dliflc.edu/mission.html>
- Dickinson, L. (1987). *Self-instruction in language learning*. (1st ed.). New York, NY: Press Syndicate of the University of Cambridge.
- Elfiky, S. A. (2012). *Investigating the relationship between students' self-assessment and ratings obtained from a formal oral proficiency interview (OPI)*.
- Eurostat. (2012). *Foreign language learning statistics*. Retrieved from http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Foreign_language_learning_statistics
- Fenstermacher, H. (2012). *Language drives economic growth, creates jobs, and fosters competitiveness for U.S. businesses*. <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Books.

- Gardner, H. (1993). *Multiple intelligences: the theory in practice*. New York, NY: Basic Books.
- Goodman, A. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- Grabe, W. (2009). *Reading in a second language, moving from theory to practice*. New York, NY: Cambridge University Press.
- Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7(2), 174-201.
- Husseinali, G. (2006). Who is studying Arabic and why? A survey of Arabic Students' Orientations at a Major University. *Foreign Language Annals*, 39(3), 395-412.
- Interagency Language Roundtable [ILR]. (2012a). Interagency Language Roundtable Language Skill Level Descriptions –Reading. Retrieved from <http://www.govtilr.org/Skills/ILRscale4.htm>
- Interagency Language Roundtable [ILR]. (2012b). *Speaking Self-Assessment*. Retrieved from <http://www.govtilr.org/Publications/speakingsa.html>
- Jackson, G. (2012). Private conversation. Research Specialist, DLIFLC, Monterey, CA.
- Jiang, B. (1999). Transfer in the academic language development of post-secondary ESL students. California State University, Fresno and University of California, Davis. *ProQuest Dissertations and Theses*. Retrieved from <http://search.proquest.com/docview/304600066?accountid=34899>. (304600066).
- Junor, L. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- Keuhl, R.O. (2000). *Design of experiments: Statistical principles of research design and analysis*. Pacific Grove, CA: Duxbury Press.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *Teachers of English Speakers and Other Languages Quarterly*, 19(4), 673-687.

- Marshall, C., & Rossman, G. B. (2008). *Designing Qualitative Research* (4th ed.). Thousand Oaks, CA: Sage Publications.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.
- Moore, D. S., & McCabe, G. P. (2006). *Introduction to the practice of statistics*. New York, NY: W. H. Freeman.
- Nordin, G. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- O'Connell, M. E., & Norwood, J. L. (2007). *International education and foreign languages, keys to securing America's future*. Retrieved from http://www.nap.edu/openbook.php?record_id=11841&page=R1
- Ochoa, E. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- Panetta, L. (2011). *Memorandum for Secretaries of the Military Departments*. United States Department of Defense. Retrieved from www.defense.gov
- Pinto, C. M. (2009). A study of seventh grade students' reading comprehension and motivation after explicit instruction in self-assessment and metacognitive reading strategies. Widener University). *ProQuest Dissertations and Theses*, 148. Retrieved from <http://search.proquest.com/docview/305136147?accountid=34899>. (305136147).
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61(2), 325-366. doi: 10.1111/j.1467-9922.2011.00640.x.
- Richards, J. (2001). *Curriculum Development in Language Teaching* New York, NY: Cambridge University Press.
- Ritchie, W. (2009). *The new handbook of second language acquisition* (2nd ed). United Kingdom: Emerald Group Publishing Limited. Howard House.
- Ross, John A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11(10), 1-13.

- Royer, D. J., & Gilles, R. (1998). Directed self-placement: an attitude of orientation. *College Composition & Communication*, 50(1), 54-70.
- Shen, H. (2002). Motivational and self-regulated learning components in relation to language learners' self-assessment, reading strategy use and reading achievement. Seattle Pacific University. *ProQuest Dissertations and Theses*, 174. Retrieved from <http://search.proquest.com/docview/305478204?accountid=34899>. (305478204).
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
- Sternberg, R. (2002). *Educational Psychology*. Boston, MA: Allyn & Bacon.
- Taha, T. (2007). Arabic as “a critical-need” Foreign Language in Post-9/11 Era: A study of students’ attitudes and motivation. *Journal of Instructional Psychology*, 34(3), 150-160.
- Thomas-Greenfield, L. (2012). *A national security crisis: foreign language capabilities in the federal government*. Retrieved from <http://www.hsgac.senate.gov/subcommittees/oversight-of-government-management/hearings/a-national-security-crisis-foreign-language-capabilities-in-the-federal-government>.
- Valencia, S. W. (2002). Understanding assessment: Putting together the puzzle. Retrieved from <http://www.eduplace.com/state/author/valencia.pdf>
- Wan-a-rom, U. (2008). Comparing the Vocabulary of Different Graded-Reading Schemes. *Reading in a Foreign Language*, 20(1), 43-69.
- Wolochuk, A. (2009). Adult English learners' self-assessment of second language proficiency: Contexts and conditions. New York University. *ProQuest Dissertations and Theses*, 169 Retrieved from <http://search.proquest.com/docview/304958045?accountid=34899>. (304958045).
- Woo, B. (1995). Interlanguage interference in adult acquisition of Korean as a second and a third language. University of San Francisco. *ProQuest Dissertations and Theses*, 189. Retrieved from <http://search.proquest.com/docview/304292420?accountid=34899>. (304292420).
- Woolfolk, A. (2007). *Educational psychology*. (10th ed.). Boston, MA: Allyn & Bacon.
- Yin, R. K. (2013). *Case study research: Design and methods*, 5th edition. Thousand Oaks, CA: Sage.

Yuko, G. B., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31.

APPENDICES

APPENDIX A**DLPT5-Reading Multiple Choice Format**

Appendix A: DLPT5-Reading Multiple Choice Format

DLPT5 in Multiple-Choice Format

Upper-Range

The Upper-Range Reading Test contains approximately 36 questions with approximately 14 authentic written passages. Each passage may have up to 5 questions with four answer choices per question.
The Upper-Range Listening Test contains approximately 36 questions with approximately 14 authentic audio passages. Each passage may have up to 3 questions with four answer choices per question. All passages are played twice.
For research purposes, some questions are not scored. These questions do not count toward the final score the examinee receives. Examinees are told that such questions are in the test but are not told which questions are the unscored ones.

DLPT5 in Multiple-Choice Format

Lower-Range

The Lower-Range Reading Test contains approximately 60 questions with approximately 36 authentic written passages. Each passage may have up to 4 questions with four answer choices per question.
The Lower-Range Listening Test contains approximately 60 questions with approximately 37 authentic audio passages. Each passage may have up to 2 questions with four answer choices per question. Passages at the beginning of the test are played once. Starting from level 2, examinees hear the passages twice.
For research purposes, some questions are not scored. These questions do not count toward the final score the examinee receives. Examinees are told that such questions are in the test but are not told which questions are the unscored ones.

APPENDIX B

DLPT5-Reading Constructed-Response Format

Appendix B: DLPT5-Reading Constructed-Response Format

DLPT5 in Constructed-Response Format

Upper-Range

The Upper-Range Reading Tests contains approximately 35 questions with 12 authentic written passages. Each passage has two or three questions.
--

The Upper-Range Listening Test contains approximately 35 questions with 12 authentic audio passages. Each passage has two or three questions and is played twice.

DLPT5 in Constructed-Response Format

Lower-Range

The Lower-Range Reading Test contains 60 questions with 30 authentic written passages. Each passage may have up to 3 questions.

The Lower-Range Listening Test contains 60 questions with 30 authentic audio passages. Each passage has two questions and is played twice.
--

APPENDIX C**Interagency Language Roundtable Language Skill Level Descriptions**

Appendix C: Interagency Language Roundtable Language Skill Level Descriptions

Reading

Preface

The following proficiency level descriptions characterize comprehension of the written language. Each of the six "base levels" implies control of any previous "base level's" functions and accuracy. The "plus level" designation will be assigned when proficiency substantially exceeds one base skill level and does not fully meet the criteria for the next "base level." The "plus level" descriptions are therefore supplementary to the "base level" descriptions. A skill level is assigned to a person through an authorized language examination.

Examiners assign a level on a variety of performance criteria exemplified in the descriptive statements. Therefore, the examples given here illustrate, but do not exhaustively describe, either the skills a person may possess or situations in which he/she may function effectively. Statements describing accuracy refer to typical stages in the development of competence in the most commonly taught languages in formal training programs. In other languages, emerging competence parallels these characterizations, but often with different details.

Unless otherwise specified, the term "native reader" refers to native readers of a standard dialect. "Well-educated," in the context of these proficiency descriptions, does not necessarily imply formal higher education. However, in cultures where formal higher education is common, the language-use abilities of persons who have had such education are considered the standard. That is, such a person meets contemporary expectations for the formal, careful style of the language, as well as a range of less formal varieties of the language.

In the following descriptions, a standard set of text-types is associated with each level. The text-type is generally characterized in each descriptive statement. The word "read," in the context of these proficiency descriptions, means that the person at a given skill level can thoroughly understand the communicative intent in the text-types described. In the usual case, the reader could be expected to make a full representation, thorough summary, or translation of the text into English. Other useful operations can be performed on written texts that do not require the ability to "read" as defined above. Examples of such tasks which people of a given skill level may reasonably be expected to perform are provided, when appropriate, in the descriptions.

R-0: Reading 0 (No Proficiency)

No practical ability to read the language. Consistently misunderstands or cannot comprehend at all.

R-0+: Reading 0+ (Memorized Proficiency)

Can recognize all the letters in the printed version of an alphabetic system and high-frequency elements of a syllabary or a character system. Able to read some or all of the following: numbers, isolated words and phrases, personal and place names, street signs, office and shop designations. The above often interpreted inaccurately. Unable to read connected prose.

R-1: Reading one (Elementary Proficiency)

Sufficient comprehension to read very simple connected written material in a form equivalent to usual printing or typescript. Can read either representations of familiar

formulaic verbal exchanges or simple language containing only the highest frequency structural patterns and vocabulary, including shared international vocabulary items and cognates (when appropriate). Able to read and understand known language elements that have been recombined in new ways to achieve different meanings at a similar level of simplicity. Texts may include descriptions of persons, places or things: and explanations of geography and government such as those simplified for tourists. Some misunderstandings possible on simple texts. Can get some main ideas and locate prominent items of professional significance in texts that are more complex. Can identify general subject matter in some authentic texts.

R-1+: Reading 1+ (Elementary Proficiency, Plus)

Sufficient comprehension to understand simple discourse in printed form for informative social purposes. Can read material such as announcements of public events, simple prose containing biographical information or narration of events, and straightforward newspaper headlines. Can guess at unfamiliar vocabulary if highly contextualized, but with difficulty in unfamiliar contexts. Can get some main ideas and locate routine information of professional significance in texts that are more complex. Can follow essential points of written discussion at an elementary level on topics in his/her special professional field. In commonly taught languages, the individual may not control the structure well. For example, basic grammatical relations are often misinterpreted, and temporal reference may rely primarily on lexical items as time indicators. Has some difficulty with the cohesive factors in discourse, such as matching pronouns with referents. May have to read materials several times for understanding.

R-2: Reading 2 (Limited Working Proficiency)

Sufficient comprehension to read simple, authentic written material in a form equivalent to usual printing or typescript on subjects within a familiar context. Able to read with some misunderstandings straightforward, familiar, factual material, but in general insufficiently experienced with the language to draw inferences directly from the linguistic aspects of the text. Can locate and understand the main ideas and details in material written for the general reader. However, persons who have professional knowledge of a subject may be able to summarize or perform sorting and locating tasks with written texts that are well beyond their general proficiency level. The individual can read uncomplicated, but authentic prose on familiar subjects that are normally presented in a predictable sequence, which aids the reader in understanding. Texts may include descriptions and narrations in contexts such as news items describing frequently occurring events, simple biographical information, social notices, formulaic business letters, and simple technical material written for the general reader. Generally, the prose that can be read by the individual is predominantly in straightforward/high-frequency sentence patterns. The individual does not have a broad active vocabulary (that is, which he/she recognizes immediately on sight), but is able to use contextual and real-world cues to understand the text. Characteristically, however, the individual is quite slow in performing such a process. Is typically able to answer factual questions about authentic texts of the types described above.

R-2+: Reading 2+ (Limited Working Proficiency, Plus)

Sufficient comprehension to understand most factual material in non-technical prose as well as some discussions on concrete topics related to special professional interests. Is markedly more proficient at reading materials on a familiar topic. Is able to separate the

main ideas and details from lesser ones and uses that distinction to advance understanding. The individual is able to use linguistic context and real-world knowledge to make sensible guesses about unfamiliar material. Has a broad active reading vocabulary. The individual is able to get the gist of main and subsidiary ideas in texts, which could only be read thoroughly by persons with much higher proficiencies. Weaknesses include slowness, uncertainty, inability to discern nuance and/or intentionally disguised meaning.

R-3: Reading 3 (General Professional Proficiency)

Able to read within a normal range of speed and with almost complete comprehension a variety of authentic prose material on unfamiliar subjects. Reading ability is not dependent on subject matter knowledge, although it is not expected that the individual can comprehend thoroughly subject matter which is highly dependent on cultural knowledge or which is outside his/her general experience and not accompanied by explanation. Text-types include news stories similar to wire service reports or international news items in major periodicals, routine correspondence, general reports, and technical material in his/her professional field; all of these may include hypothesis, argumentation and supported opinions. Misreading rare. Usually able to interpret material correctly, relate ideas and "read between the lines," (that is, understand the writers' implicit intents in text of the above types). Can get the gist of more sophisticated texts, but may be unable to detect or understand subtlety and nuance. Rarely has to pause over or reread general vocabulary. However, may have trouble with unusually complex structure and low frequency idioms.

R-3+: Reading 3+ (General Professional Proficiency, Plus)

Can comprehend a variety of styles and forms pertinent to professional needs. Rarely misinterprets such texts or rarely experiences difficulty relating ideas or making inferences. Able to comprehend many sociolinguistic and cultural references. However, may miss some nuances and subtleties. Able to comprehend a considerable range of intentionally complex structures, low frequency idioms, and uncommon connotative intentions, however, accuracy is not complete. The individual is typically able to read with facility, understand, and appreciate contemporary expository, technical or literary texts, which do not rely heavily on slang and unusual items.

R-4: Reading 4 (Advanced Professional Proficiency)

Able to read fluently and accurately all styles and forms of the language pertinent to professional needs. The individual's experience with the written language is extensive enough that he/she is able to relate inferences in the text to real-world knowledge and understand almost all sociolinguistic and cultural references. Able to "read beyond the lines" (that is, to understand the full ramifications of texts as they are situated in the wider cultural, political, or social environment). Able to read and understand the intent of writers' use of nuance and subtlety. The individual can discern relationships among sophisticated written materials in the context of broad experience. Can follow unpredictable turn of thought readily in, for example, editorial, conjectural, and literary texts in any subject matter area directed to the general reader. Can read essentially all materials in his/her special field, including official and professional documents and correspondence. Recognizes all professionally relevant vocabulary known to the educated non-professional native, although may have some difficulty with slang. Can read reasonably legible handwriting without difficulty. Accuracy is often nearly that of a

well-educated native reader.

R-4+: Reading 4+ (Advanced Professional Proficiency, Plus)

Nearly native ability to read and understand extremely difficult or abstract prose, a very wide variety of vocabulary, idioms, colloquialisms and slang. Strong sensitivity to and understanding of sociolinguistic and cultural references. Little difficulty in reading less than fully legible handwriting. Broad ability to "read beyond the lines" (that is, to understand the full ramifications of texts as they are situated in the wider cultural, political, or social environment) is nearly that of a well-read or well-educated native reader. Accuracy is close to that of the well-educated native reader, but not equivalent.

R-5: Reading 5 (Functionally Native Proficiency)

Reading proficiency is functionally equivalent to that of the well-educated native reader. Can read extremely difficult and abstract prose; for example, general legal and technical as well as highly colloquial writings. Able to read literary texts, typically including contemporary avant-garde prose, poetry and theatrical writing. Can read classical/archaic forms of literature with the same degree of facility as the well-educated, but non-specialist native. Reads and understands a wide variety of vocabulary and idioms, colloquialisms, slang, and pertinent cultural references. With varying degrees of difficulty, can read all kinds of handwritten documents. Accuracy of comprehension is equivalent to that of a well-educated native reader.

APPENDIX D

Self-Assessment Survey of Reading Proficiency

Appendix D: Self-Assessment Survey of Reading Proficiency

Reading Levels:

level 0+	Statements 1-2
level 1	Statements 3-9
Level 1+	Statements 10-17
Level 2	Statements 18-24
Level 2+	Statements 25-34
Level 3	Statements 35-42

APPENDIX E

Can-Do-Scale Instrument

Appendix E: Can-Do-Scale Instrument

Self-Assessment Survey of Reading Proficiency Can-Do-Scale (CDS)

The following self-assessment of Reading ability is intended to guide those who have not taken a U.S government-sponsored Arabic DLPT-5 Reading test. It will produce an estimate of your Arabic Reading ability but is in no way a replacement for a formal Defense Language Proficiency Test 5 Reading skill (DLPT5-R).

Important: The term *read* as used in this self-assessment always means “*read and understand the meaning.*” It does not refer in any way to the ability to read aloud without comprehension. The term *text* refers to any example of language presented in the writing system of the Arabic language, including advertisements, weather reports, news articles, letters, lengthy essays, and literary works, among others.

PLEASE PRINT

Name: Last _____ First _____

Class number: _____

- 1) Language tested: Arabic
- 2) Education Level Completed:

Non-high school graduate - GED - HS - AA - BA - MA
(Circle the highest level you have completed.)

- 3) Military branch: Army- Air Force- Navy- Marines (Circle one)

- 4) Rank: A) Officer B) Enlisted (Circle one)

- 5) Gender: Male – Female (Circle one)

- 6) Age: A) (18- 20) B) (21-25) C) (26-30)
D) (31-35) E) (36- 40) (Circle one)

- 7) Did you grow up reading a language other than English? A) Yes B) No

If your answer is B) (**No**), please go on to question 8.

- a. If yes, which language? _____
- b. If yes, at what age did you start reading English? _____

- 8) Have you studied Arabic before coming to the DLIFLC? A) Yes B) No

- a. If yes, for how long have you studied Arabic? _____ Years,
_____ Months.

9) Have you studied any other foreign language(s)?

A) No B) Yes

- a) If yes, which language(s) and for how long?

#	Name of the language	Years	Months
1			
2			
3			
4			

Below is a series of statements describing reading tasks that require use of a foreign language. Please read each statement carefully and check the appropriate box to indicate how well you can perform the task in Arabic.

Please complete the survey by answering the 42 items.

<p>1- In reading Arabic, I can recognize and identify all the letters of the Arabic alphabet and the elements of the Arabic writing system. (e.g., Arabic is written from right to left, some Arabic letters do not connect to letters that follow).</p> <p> <input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all </p>
<p>2- I can read some isolated Arabic words and phrases, such as numbers and commonplace names that I see on signs, menus, and storefronts, and in simple everyday material such as advertisements and timetables.</p> <p> <input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all </p>
<p>3- I can understand the purpose and main idea of short and simple Arabic texts, such as in printed business advertisements, public announcements, maps, etc.</p> <p> <input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all </p>
<p>4- I can understand simple instructions in Arabic text, such as in straightforward street directions, or</p>

<p>written directions to where an apartment is located.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>5- I can understand short, simple descriptions written in Arabic, of familiar persons, places, and things, like those found in many tourist pamphlets.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>6- I can understand simple Arabic texts related to social or practical activities, e.g., personal invitations to a marriage party, engagement party, or birthday party.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>7- I can understand simple Arabic texts related to travel regulations (e.g., lost luggage, prohibited items, weight limit).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>8- I can understand simple congratulatory messages written in Arabic (e.g., about having a new-born baby, high school graduation, buying a new house).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>9- I can read and understand simple Arabic reading texts that use cognates, the most generic and most common vocabulary, e.g., about geographical features and climate.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>10- I can read and understand Arabic material such as announcements of public events.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>11- I can read simple Arabic text containing biographical information or narration of events.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>12- I can read and understand straightforward Arabic newspaper headlines.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>13- I can get some main ideas and locate routine information of professional significance in more complex Arabic texts (e.g., about education in the Arab world).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>14- I can understand the main ideas and simple explicitly stated information in Arabic reading texts on topics with which I am familiar (e.g., about entertainment activities, including film festivals and television series).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>

<p>15- I can understand “who, what, when, or where” in Arabic texts that are predictable in their content.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>16- I can understand simple, straightforward paragraphs in Arabic that contain descriptive information about concrete things like a house or a person.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>17- In reading simple, straightforward paragraphs in Arabic, I can guess the meaning of unfamiliar vocabulary from context with which I am familiar (e.g., hobbies, sports).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>18- I can understand Arabic texts that consist mainly of straightforward factual language, such as short news reports of events, biographical information, descriptions, or simple technical material.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>19- I can understand the main idea and some details of clearly organized short, straightforward Arabic texts about places, people, and events with which I am familiar.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>20- I can understand straightforward Arabic reports about current and past events.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>21- I can understand simple letters in Arabic on familiar topics, including descriptions of friends’ current activities and future plans.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>22- I can usually understand the main ideas of Arabic authentic texts on topics with which I am familiar, either because they pertain to my work experience or to topics in which I am interested (e.g., natural disasters and environment).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>23- I can use basic Arabic cultural knowledge to help me interpret concrete information that I read in Arabic texts (e.g., Natural resources and border conflicts in the Middle East).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>

<p>24- I can follow the development of events described in Arabic texts, such as the sequence of events, including cause and effect relationship (e.g., unemployment and poverty).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>25- I can understand the main idea and major details of Arabic texts that are not necessarily presented in a predictable or straightforward way.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>26- I can follow the development of events mentioned in Arabic texts dealing with concrete and some abstract topics about common societal issues (e.g., marriage of children under 18).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>27- I can draw simple inferences or conclusions based on factual information presented in an Arabic text (e.g., military maneuvers that threaten countries).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>28- I can understand most factual material in non-technical Arabic as well as some discussions on concrete topics (e.g., Iran's military inventions and imposed economic sanctions).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>29-In reading Arabic, I can separate the main idea and details from lesser ones and use that distinction to advance understanding.</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>30-In reading Arabic, I can use linguistic context and real-world knowledge to make sensible guesses about the meaning of unfamiliar material (e.g., about trading of human organs).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>31-I can understand most factual reading passages from a magazine (e.g., fast food restaurants and their positive and negative effects on family relationships and human health).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>32-I can understand main and supporting ideas in reading passages (e.g., family violence against women, its cultural causes and its effects).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>33-I can understand the main and supporting ideas in reading passages about economic issues in the Middle East (e.g., a discussion about an increase in food prices and future expectations).</p> <p><input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all</p>
<p>34-I can understand the main and supporting ideas in reading passages from newspapers about science and technology (e.g., scientific and medical inventions, like organ transplants and smart phone applications).</p>

<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
35- I can usually read Arabic and understand all of the material in a major daily newspaper published in a city or country with which I am familiar.
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
36- In reading an Arabic newspaper or magazine that contains editorial or opinion content, I can “read between the lines” and understand meanings that are not directly stated.
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
37- I can understand the author’s intent and follow the line of reasoning in Arabic texts that include persuasion, supported opinion or argument for a position (e.g., editorials, debates, and op-ed pieces) with little or no use of a dictionary.
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
38- I can understand the main idea and important details of almost all Arabic material written within my particular professional field or area of primary interest (e.g., military reports, medical reports, historical reports, etc.).
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
39- I can understand Arabic texts that include argumentation and supported opinions, like those found on the internet about international news (e. g., newspapers, magazines, and journals).
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
40- In reading Arabic, I can understand comparisons of points of view (e.g., opinion pieces and political analyses).
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
41- I can almost completely understand a variety of Arabic materials on unfamiliar subjects (e.g., combating corruption and bribery in the Arab world, and religions dialogues to combat extremism).
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all
42- In reading Arabic, I can always interpret material correctly and misreading is rare (e.g., general international news from the internet).
<input type="checkbox"/> Quite Easily <input type="checkbox"/> Easily <input type="checkbox"/> With some difficulty <input type="checkbox"/> With great difficulty <input type="checkbox"/> Not at all

Thank you

APPENDIX F

Informed Consent Form

Appendix F: Informed Consent Form

TITLE OF STUDY

INVESTIGATING THE RELATIONSHIP BETWEEN STUDENTS' SELF-ASSESSMENT AND RATINGS OBTAINED FROM A FORMAL DEFENSE LANGUAGE PROFICIENCY TEST 5 READING SKILLS (DLPT5-R)

RESEARCHER

Mr. Mohamad Alkhatatbeh

PURPOSE OF THE STUDY

The purpose of the study is to investigate the relationship between a self-assessment instrument of Arabic reading ability and a formal Arabic DLPT5-Reading scores.

This study as a part of doctoral research in Instructional Leadership at Argosy University, San Francisco.

If you voluntarily decide to participate in the study, you will complete the self-assessment instrument (survey) about your Arabic reading skills. Then, your scores on the instrument will be compared with your formal Arabic DLPT5-Reading results. As your scores on the Oral Proficiency Interview (OPI), on DLPT5 for Reading and Listening will be requested from DLIFLC for future studies

CONFIDENTIALITY

The information that will be gathered will be maintained with high confidentiality and applicable to all federal regulation and laws. The limits of confidentiality may be broken in the case of a court subpoena, or other lawful means. To maintain student anonymity, you will be assigned a participant number instead of using your name or any information that may identify you. The information gathered about you will be kept safe and saved in a safety box of a bank. After five years from completing the study all the survey documents that you completed in will be destroyed.

WHAT YOU WILL DO

If you agree to participate in the study, you will complete self-assessment questionnaire about your Arabic Reading skill. First, you will complete biographical information about you to examine its relationship with accuracy of self-assessment. Second, you will specify how easy/difficult to read Arabic texts. Completing the survey will take about 40 minutes, and it will administrate before one-to-two week of your DLPT5 test.

RISKS AND BENEFITS

Participation in the study is voluntary and there will be no negative penalties or any effects if you decide not to participate, and you may withdraw from the study at any time. You will be able to get your survey score for your future language studies from the researcher.

PLEASE NOTE THE FOLLOWING

If you experience any level of discomfort during the survey, you have the option of skipping the discomfort-causing question or ending the survey all together; you will be asked to contact your Military Unit for follow up. A California certified mental health professional with a telephone number will be recommended and provided to you at any point you need it, during or after completing the survey; (Dr. Tabije, Jon R: Presidio of Monterey U.S. Army Behavioral Health Clinic; M, T, W, F 0730-1630, and TH 0730-1200, Presidio of Monterey, DLIFLC, 473 Cabrillo St. Bldg. 422, Monterey, CA 93944

(831) 242-4328. Dr. Tabije, Jon R is available and reachable, as he is listed as a staff person of DLI and is listed on The Presidio of Monterey U.S. Army Health Clinic roster).

MORE INFORMATION

If you have any questions or would like more information, please call me at 831-242-6940, send me an email at Mohaahmad1@yahoo.com, or stop by my office at the POM, Bldg. 635 B Room 6 in Monterey California. You may also contact my dissertation advisor at Argosy University, Dr. Quamina Afriye, via email at Afriyequamina@mac.com.

I have read the above and I understand its contents and I agree to participate in the study.

PLEASE PRINT

Last name	First name	Middle initial	Date
_____ Signature			

APPENDIX G

The Approval Letter to Conduct the Study at DLIFLC

Appendix G: The Approval Letter to Conduct the Study at DLIFLC



DEPARTMENT OF THE ARMY
DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER
AND PRESIDIO OF MONTEREY
MONTEREY CA 93944-3236

October 1, 2013

Institutional Review Board (IRB)
U.S. Army Assurance: DOD A20209

Sylvie Merchant, Ph.D.
IRB Chair
Argosy University
1005 Atlantic Ave.
Alameda, CA 94501

Dear Dr. Merchant:

On behalf of the U.S. Army Defense Language Institute Foreign Language Center (DLIFLC), I am writing to formally indicate our awareness of a research project proposed by Mr. Mohamad Alkhatabeh, a graduate student (School of Education) at Argosy University.

This research project, tentatively entitled *Investigating the Relationship Between Students' Self-Assessment and Ratings Obtained From a Formal Defense Language Proficiency Test 5 Reading Skills (DLPTS-R)* has been reviewed and approved by the DLIFLC Scientific & Ethics Review Boards, by Dr. Jielu Zhao (DLIFLC Associate Provost for Undergraduate Education) and by Dr. Marina Cobb (DLIFLC Dean, Middle East Languages III). Dr. Zhao has endorsed the use of DoD military personnel as participants in Mr. Alkhatabeh's dissertation research project.

I have been informed that the Argosy IRB will conduct the review and maintain institutional oversight of this project. Once the Argosy IRB has completed its review of the project, I ask that a copy of the outcome of that review (date, review type, & approval number) be sent to me so we may maintain a folder on this project in our file of current research projects.

If you have any questions or concerns, please feel free to contact me.

Sincerely,

J. Jeffrey Crowson, Ph.D.
IRB Chair
Professor, Educational Research
(831) 242-3788
jeffrey.j.crowson.civ@mail.mil